

Sequence alignment visualization in HTML5 without Java

Christoph Gille^{1,*}, Weyand Birgit² and Andreas Gille^{3,†}

¹Universitätsmedizin Berlin, Charité, Institut für Biochemie, 10117 Berlin, Germany, ²Department of Plastic, Hannover Medical School, Hand and Reconstructive Surgery, 30625 Hannover, Germany and ³Universität Heidelberg, Institut für Experimentelle und Klinische Pharmakologie und Toxikologie, 68169 Mannheim, Germany

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Java has been extensively used for the visualization of biological data in the web. However, the Java runtime environment is an additional layer of software with an own set of technical problems and security risks. HTML in its new version 5 provides features that for some tasks may render Java unnecessary.

Results: ALIGNMENT-TO-HTML is the first HTML-based interactive visualization for annotated multiple sequence alignments. The server side script interpreter can perform all tasks like (i) sequence retrieval, (ii) alignment computation, (iii) rendering, (iv) identification of a homologous structural models and (v) communication with BioDAS-servers. The rendered alignment can be included in web pages and is displayed in all browsers on all platforms including touch screen tablets. The functionality of the user interface is similar to legacy Java applets and includes color schemes, highlighting of conserved and variable alignment positions, row reordering by drag and drop, interlinked 3D visualization and sequence groups. Novel features are (i) support for multiple overlapping residue annotations, such as chemical modifications, single nucleotide polymorphisms and mutations, (ii) mechanisms to quickly hide residue annotations, (iii) export to MS-Word and (iv) sequence icons.

Conclusion: ALIGNMENT-TO-HTML, the first interactive alignment visualization that runs in web browsers without additional software, confirms that to some extent HTML5 is already sufficient to display complex biological data. The low speed at which programs are executed in browsers is still the main obstacle. Nevertheless, we envision an increased use of HTML and JavaScript for interactive biological software.

Availability and implementation: Under GPL at: <http://www.bioinformatics.org/strap/toHTML/>.

Contact: christoph.gille@charite.de

Received on September 10, 2013; revised on October 11, 2013; accepted on October 19, 2013

1 INTRODUCTION

The amount of experimental sequence-related data, such as post-translational modifications, ligand binding and genetic variants is rapidly growing. Efficient web-based visualization facilitates understanding of the biological meaning of comprehensive data sets. Advanced visualization of multiple sequence alignments (MSA) includes sequence-related information such as residue annotation, color-coding of physical properties of residues,

secondary structure and conservation/variation. A number of programs have been developed for non-interactive rendering of MSAs: Alscript (Barton, 1993), Boxshade, Esprint (Gouet *et al.*, 2003) and TexShade (Beitz, 2000). These programs generate an alignment figure according to a script file. The script can be prepared manually in a text editor or dynamically by another program. Therefore, these programs are suited to display computational results of web services. The generated alignment figures use vector graphics formats for quality reasons, but most browsers do not support vector image formats (e.g. SVG) and external document viewers are required. A further disadvantage is that these figures are static and lack interactivity. Interactive capabilities, hyperlinks, balloon messages and the option to shield information become increasingly important in the omics era.

Java brought interactive functionality to web pages and allows alignment manipulation in real time (Caffrey *et al.*, 2007; Clamp *et al.*, 2004; Lord *et al.*, 2002). Owing to its cross-platform compatibility, integrated graphics and high execution speed, Java is widely used.

But there are also five potential problems: (i) failure to start Java applications, (ii) security violations, (iii) browser crashes, (iv) incompatibility on some platforms such as Android OS and (v) slowing of the computer. Today the web standard HTML5 is being increasingly used for interactive web applications such as genome browsers (Hoy, 2011; Medina *et al.*, 2013; Miller *et al.*, 2013). The main advantage is that HTML5 code is executed directly in the browsers without additional client side software. A potential challenge with HTML5 code is that its execution is slow, and we were wondering whether it would be sufficient for full-featured MSA visualization.

2 IMPLEMENTATION

Our alignment visualization system consists of two parts: (i) A server side non-interactive program generates a graphical representation of the MSA in HTML according to an input text file. (ii) A comprehensive JS-library encoding the interactive graphical user interface for the HTML page.

Canvas elements allow for dynamic scriptable rendering of 2D shapes and are used for secondary structure cartoons. If they are not supported, the secondary structure is symbolized by text characters. For maximum cross-browser compatibility, residues are drawn as classical text elements with specific color, background and style attributes. This approach is robust and independent of JavaScript. We found that client side rendering often takes several seconds for large alignments. This is caused by the

*To whom correspondence should be addressed.

†Present address: CSL Limited, 45 Poplar Road, 3052 Parkville, Australia.

low speed of JavaScript and manipulation of the document object model. To prevent the initial delay when the page is loaded, MSA rendering is already performed at server side. Only if the user changes the color scheme or the conservation threshold, client side computation of sequence conservation at all positions is carried out and each residue is wrapped in a separate node. Overlapping residue underlinings are laid out efficiently such that the least amount of space below the sequence line is occupied.

3 RESULTS

ALIGNMENT-To-HTML is the first alignment visualization program for decorated MSAs using HTML featuring wrapped and one-line sequence view, rendering styles, secondary structure, drag and drop, pattern highlighting and several export options (Fig. 1). Visualization is independent on plugins and therefore robust while providing a rich user interface. Compared with existing software for alignment visualization, ALIGNMENT-To-HTML provides novel functionality: It has the ability to display multiple overlapping residue annotations and the option to export the alignment directly to MS-Word and full-featured alignment workbenches. Icons may symbolize species or biological tissues. Underlining all BioDAS sequence features for a UNIPROT entry may result in an overcrowded confusing alignment figure. Therefore check boxes and the trash bin enable reversible hiding of sequence features. Although existing alignment rendering software like Alscript, Esript, Boxshade, TexShade do not take care of alignment computation themselves, ALIGNMENT-To-HTML includes the entire logic for alignment generation, rendering and annotation. This involves data loading, alignment computation, 3D superposition, identification of homologous 3D structures and loading of annotations from DAS servers. Currently, the scripting language comprises >40 commands and numerous sequence and 3D rendering styles.

4 CONCLUSION

HTML5 is sufficient for interactive alignment visualization. It is robust because it runs directly in standard web browsers without plugins. Java is still required for 3D visualization.



Fig. 1. Example of an interactive alignment view. Only a section of the scrollable alignment is visible. The script consists of only three commands to load the PFAM entry, to associate 3D structures and to load annotations from DAS services. An asparagine residue that has been clicked in the alignment is selected in the 3D view: Some sequence annotations are deactivated with check boxes or with the trash bin such as mature protein region (MPR)

ALIGNMENT-To-HTML can be used for bioinformatics web services or for locally installed programs to generate HTML files. It takes care of all computations toward a full-rendered interactive MSA.

Conflict of Interest: none declared.

REFERENCES

- Barton,G.J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.*, **6**, 37–40.
- Beitz,E. (2000) TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics*, **16**, 135–139.
- Caffrey,D.R. et al. (2007) PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments. *BMC Bioinformatics*, **8**, 381.
- Clamp,M. et al. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.
- Gouet,P. et al. (2003) ESript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.*, **31**, 3320–3323.
- Hoy,M.B. (2011) HTML5: a new standard for the Web. *Med. Ref. Serv. Q.*, **30**, 50–55.
- Lord,P.W. et al. (2002) CINEMA-MX: a modular multiple alignment editor. *Bioinformatics*, **18**, 1402–1403.
- Medina,I. et al. (2013) Genome Maps, a new generation genome browser. *Nucleic Acids Res.*, **41**, 41–46.
- Miller,C.A. et al. (2013) Scribl: an HTML5 Canvas-based graphics library for visualizing genomic data over the web. *Bioinformatics*, **29**, 381–383.