WAViS server for handling, visualization and presentation of multiple alignments of nucleotide or amino acids sequences

Radek Zika, Jan Pačes, Adam Pavlíček and Václav Pačes*

Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, CZ-16637 Prague, Czech Republic and Institute of Chemical Technology Prague, Czech Republic

ABSTRACT

Web Alignment Visualization Server contains a set of web-tools designed for quick generation of publication-quality color figures of multiple alignments of nucleotide or amino acids sequences. It can be used for identification of conserved regions and gaps within many sequences using only common web browsers. The server is accessible at http://wavis.img.cas.cz.

INTRODUCTION

Contemporary genomics requires comparisons and alignments of nucleotide or amino acids sequences generated in genomics projects. The multiple alignments are often based on comparing sequences that contain many small and large non-matching regions, indels and gaps of varying lengths. There are several utilities for graphical presentation of alignments, such as TeX-shade (1) and Pfaat (2). These programs require local installations. Other programs, such as ESPript (3) and BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html) have a web interface.

We describe here a new server (Web Alignment Visualization Server, WAViS) that makes it easy to prepare graphical outputs of the multiple alignments in a platform-independent vector format. These outputs are in modes ready for presentations and for further modifications in professional drawing programs (e.g. Adobe Photoshop or CorelDraw).

THE FEATURES OF WAViS

The server program components are written in the PERL language using GD (http://www.boutell.com/gd) and CGI libraries (http://stein.cshl.org/www/software/cgi). Standalone scripts are written for the UNIX (in particular for the Linux)

Table 1. Input and output formats and their suffixes used by WAViS

Input Format	Common suffix(es)	Output Format	Common suffix(es)
multi-FASTA	mfa	Portable network graphics	png
NEXUS	nex	Encapsulated postscript	eps, eps.gz, eps.bz2
GCG	msf	Scalable vector graphics	svg, svg.gz, svg.bz2

PostScript: http://www.postscript.org/FAQs/language/FAQ.html; SVG: http://www.w3.org/Graphics/SVG.

environment. Various formats of multiple alignment files can be converted to a multi-FASTA alignment (MFA) file using a simple conversion utility. Its output is stored and subsequently used for the drawing routines.

The web interface allows the user to upload files and to generate color bitmap files of aligned sequences. The user can obtain a preview picture, select and set up options and upload data. The output is in SVG (vector) format and it can be converted to bitmap formats. The input has to be in one of the following formats: MFA, Clustal (4), NEXUS (5), GCG/MSF (http://www.gcg.com). The bitmap outputs can be imported into documents suitable for presentation. In addition the encapsulated postscript can be compressed. The user can choose between two alternative compression methods.

The picture's dimensions and other properties can be defined in the input form, which also allows the sequences to be sorted by several characteristics. For any part of the sequence, a specific color can be predefined. Another possibility is to produce alignment graphics in the vector SVG format. This format is suitable for further handling in specialized graphical software.

*To whom correspondence should be addressed. Tel: +420 220183541; Fax: +420 224311019; Email: vpaces@img.cas.cz
Present address:
Adam Pavlíček, Genetic Information Research Institute, 2081 Landings Drive, Mountain View, CA 94043, USA.
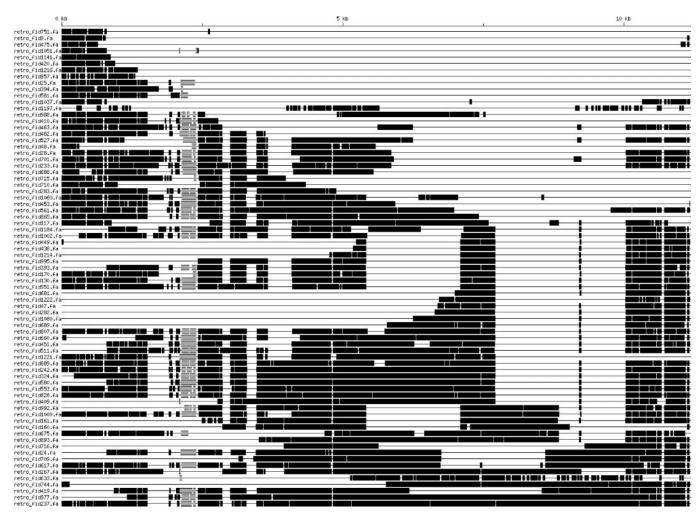
Figure 1. Alignment of human endogenous retroviral elements of the HERV-K family. The figure was prepared using data from the HERVd database (6) with the following parameters: picture width = 800, sequence width = 8.

The input and output formats used by WAViS are in Table 1.

An example of one of the possible WAViS outputs is given in Figure 1.

ACKNOWLEDGEMENTS

REFERENCES

1. Beitz,E. (2000) TeXshade: shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics*, **16**, 135–139.

2. Johnson,J.M., Mason,K., Moallemi,C., Xi,H., Somaroo,S. and Huang,E.S. (2003) Protein family annotation in a multiple alignment viewer. *Bioinformatics*, **19**, 544–545.

3. Gouet,P., Courcelle,E., Stuart,D.I. and Metoz,F. (1999) ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.

4. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22**, 4673–4680.

5. Maddison,D.R., Swofford,D.L. and Maddison,W.P. (1997) NEXUS: an extendable file format for systematic information. *Syst. Biol*., **46**, 590–621.

6. Pačes,J., Pavlíček,A., Zika,R., Kapitonov,V.V., Jurka,J. and Pačes,V. (2004) HERVd: the Human Endogenous RetroViruses Database: update. *Nucleic Acids Res*., **32**, D50.