

## Dotlet: diagonal plots in a Web browser

Thomas Junier and Marco Pagni\*

Swiss Institute of Bioinformatics, Swiss Institute for Experimental Cancer Research,  
Chemin des Boveresses 155, 1066 Epalinges, Switzerland

Received on August 18, 1999; revised on September 17, 1999; accepted on September 22, 1999

### Abstract

**Summary:** *Dotlet* is a program for comparing sequences by the diagonal plot method. It is designed to be platform-independent and to run in a Web browser, thus enabling the majority of researchers to use it.

**Availability:** The applet can be tested at <http://www.isrec.isb-sib.ch/java/dotlet/Dotlet.html>, and the source code is available upon request.

**Contact:** (Thomas.Junier|Marco.Pagni)@isrec.unil.ch

**Supplementary information:** The full documentation about *dotlet* is available from the above URL.

### Yet another diagonal plot tool?

The diagonal plot method (Gibbs and McIntyre, 1970) allows the user to compare two sequences and find regions of similarity. Its advantage over text-based alignment programs is that it allows a visual detection of repeats, inversions, regions of low complexity, and introns, among others. On the other hand, it can be less sensitive because only high-scoring segment pairs, but not gapped alignments, are evaluated.

Diagonal plots have existed for a long time, but the early versions [e.g. the GCG package *DotPlot* (Devereux *et al.*, 1984)] were rather awkward for anyone not used to a command-line interface. Often the signal-to-noise ratio could not be changed without recomputing the whole plot, which meant that several runs had to be performed until the significant features stood out clearly against the background; and the similarity was generally expressed as either a black or a white pixel.

This changed with the advent of GUI-driven versions like *Dotter* (Sonnhammer and Durbin, 1995). In these more recent implementations, the similarities are expressed by a grey scale, and the parameters can be set interactively in real time, making it easy to tune them so as to achieve a maximal signal-to-noise ratio.

However, to our knowledge there is no diagonal plot tool that runs on any operating system, hence the need for a new version that would run on as many platforms as possible. Given the ever increasing popularity of Web-

based tools in molecular biology, an applet that would run in any Web browser was an obvious choice.

### Features

*Dotlet* is fully mouse and menu driven. Sequences can be imported by cut-and-paste (there is no built-in limit to the number of sequences that can be imported), and menus allow the user to select which two sequences to compare. The score matrix, zoom factor and comparison window size are selected through menus. Several scoring matrices are available, including the BLOSUM (Henikoff and Henikoff, 1992), Gonnet (Gonnet *et al.*, 1992) and PAM (Dayhoff and Eck, 1968) series. After the user clicks the compute button, the diagonal plot is shown in the main panel (Figure 1A), which can be scrolled if the sequences are too long.

*Dotlet* supports protein versus protein, DNA versus DNA, and mixed comparisons. In the latter case, the DNA is first translated in the three forward reading frames, and a protein versus protein comparison is performed. Whether a sequence is protein or nucleic acid is automatically detected.

A histogram of score frequencies (Figure 1B) allows the user to discriminate between the noise and the regions of similarity, and to accordingly set the grey scale in such a way that all significant regions, and only those, are highlighted. The histogram is in both linear and semi-logarithmic scales, the latter being useful when there are few high-scoring pixels. Any change in the grey scale settings is reflected in the main panel in real time.

A cursor in the main panel allows the user to select a pixel in the diagonal plot and to examine the sequences at the corresponding positions, as displayed in the alignment panel (Figure 1C). Positively scoring residue pairs are shaded in blue. For nucleic acids versus protein comparisons, this panel shows the three reading frames of the DNA sequence (as protein) against the protein sequence. This makes it easy to identify introns and frame shifts. For nucleic acid versus nucleic acid comparisons, the panel shows a second alignment: the horizontal sequence against the reversed, complemented vertical sequence. The cursor can also be moved along the horizontal and vertical sequences using the scrollbars, or with the keyboard.

\*To whom correspondence should be addressed.

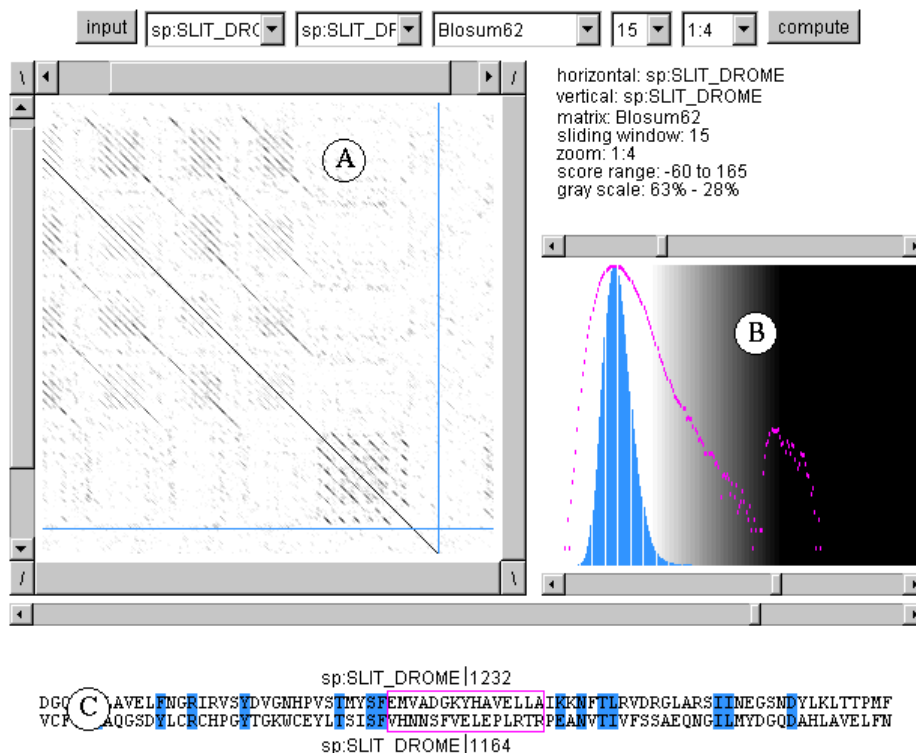


Fig. 1.

### Acknowledgements

This work was supported by grant 31-49669.96 from the Swiss National Research Foundation.

### References

- Dayhoff, M.O. and Eck, R.V. (1968) *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Silver Spring.
- Devereux, J., Haerberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.*,

12, 397–395.

- Gibbs, A.J. and McIntyre, G.A. (1970) The diagram, a method for comparing sequences. *Eur. J. Biochem.*, 16, 1–11.
- Gonnet, G.H., Cohen, M.A. and Benner, S.A. (1992) Exhaustive matching of the entire protein sequence database. *Science*, 256, 1443–1445.
- Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution from protein blocks. *Proc. Natl Acad. Sci. USA*, 98, 10915–10919.
- Sonnhammer, E.L.L. and Durbin, R. (1995) A dot matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, 167, GC1–10.