CINEMA-MX:- A modular multiple alignment editor

P.W. Lord [*], J.N. Selley, and T.K. Attwood
School of Biological Sciences
University of Manchester
Oxford Road
Manchester
M13 9PL
p.lord@russet.org.uk
[attwood,selley]@bioinf.man.ac.uk

Abstract

Summary: Analysing and visualising multiple sequence alignments is a common task in many areas of molecular biology and bioinformatics. Many tools exist for this purpose, but are not easily customisable for specific in-house uses. Here we report the development of an editor, CINEMA-MX, that addresses these issues. CINEMA-MX is highly modular and configurable, and we present examples to illustrate its extensibility.
Availability: The program and full source code, which are available from `http://www.bioinf.man.ac.uk/dbbrowser/cinema-mx`, are being released under a combination of the LGPL and GPL, for Unix or Windows platforms.
Contact: `p.lord@russet.org.uk`

Multiple alignments are important tools in sequence analysis. By aligning sequences, it is possible to determine the most similar regions, which can provide functional information about proteins by inferring homology. Consequently, alignments underpin several databases that are used to determine membership of functionally related families.

Given the known limitations of automated tools, manual editors are often used in the process of alignment visualisation and refinement (e.g., STRAP (Gille & Frömmel, 2001), MASE (Faulkner & Jurka, 1988), SEAVIEW (Galtier *et al.*, 1996)). As the uses of alignments are widespread, it is clear that the requirements of editors differ between user groups. Addressing these needs is a challenging task because an editor is likely either to lack the functionality required by some users, or to become overly large and complex, which can complicate the user interface. This situation is compounded because some existing editors are not readily extensible and/or may rely on old technology, such as Java 1.0 (e.g., CINEMA (Parrry-Smith *et al.*, 1998), Jalview `http://www.ebi.ac.uk/~michele/jalview`).

To address some of these problems, we have developed CINEMA-MX (Colour INteractive Editor for Multiple Alignments-Modular, eXtensible), using Java 2 and the Swing GUI toolkit. The application is built from ~ 20 modules, each being defined by a lightweight class. We call this the "MX" architecture.

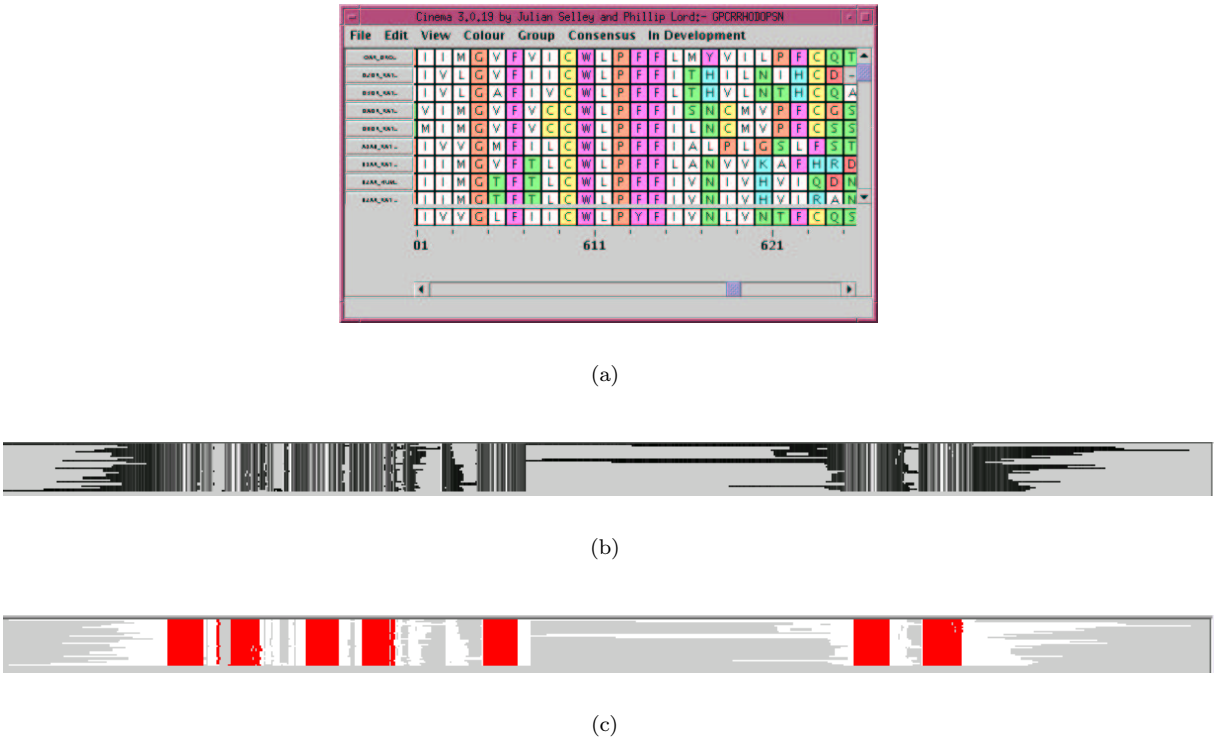
CINEMA-MX uses XML files to determine loading and to configure its modules on a per user basis as required. This has advantages over Beans components, as modules can be reconfigured without a complex development environment. Its modular architecture would allow it to be easily adapted to component integration systems, such as ISYS (Siepel *et al.*, 2001).

On loading, CINEMA-MX presents the user with a graphical display of a user-defined alignment, as shown in Figure 1(a). Gaps can be added or removed by mouse drags. Information about the sequence (such as the accession number or database identifier) is also presented.

Sequences may be coloured by a variety of different schemes, for example on a per residue basis or according to variability in alignment columns (see Figure 1(b)). New colour schemes can be "plugged in" using the configuration files.

Regions of interest or conserved motifs can be selected by mouse drags. The details of each selection can be viewed, and parameters such as the size of each selection can be modified precisely using the "motif manager" dialog.

[*]Present Address:- Department of Computing Science, University of Manchester, Oxford Road, Manchester, M13 9PL



(a)

(b)

(c)

Fig. 1. Multiple views of CINEMA-MX, showing: 1(a) the start-up window with an alignment coloured by residue properties; 1(b) an overview of the entire alignment highlighting the most conserved regions (lighter shades), which in this case are transmembrane (TM) domains; 1(c) the location of PRINTS motifs displayed on the alignment (darker shades), pin-pointing the TM domains highlighted in 1(b).

The display software is flexible. For example, the size of each residue can be shrunk down to a single pixel. Combined with the variance display, this allows an overview of the alignment, and its conservation (as in Figure 1(b)).

CINEMA-MX has a "group" facility that allows editing of, or analyses to be performed over, the sequences within a group. The aligned sequences, or their consensuses, can be viewed in a separate window. This allows display of the variability within many overlapping families and sub-families from the main alignment.

We have exploited the customisability of CINEMA-MX in a number of ways. For example, the location of PRINTS motifs (Attwood *et al.*, 2002) can be viewed (Figure 1(c)), or the EMBOSS "seqret" application can directly read files in any supported format. These serve as prototypes for further integration of the EMBOSS programs (Rice *et al.*, 2000), and for more complex queries against PRINTS-S, or other databases.

A number of further developments are planned. For example, although the editor can display large alignments (up to ~6Mb), data are stored in memory, and we would like to provide "paging". We are also integrating components, such as 3D structure and phylogenetic tree viewers, and have prototyped a version for viewing RNA alignments and their secondary structure. The MX architecture allows this application to share most of the main CINEMA-MX code base.

Acknowledgments

We are grateful for support from Pfizer Ltd (PL), and the European Molecular Biology Network (JS, PL). TKA is a Royal Society University Research Fellow.

References

Attwood, T. K., Blythe, M., Flower, D. R., Gaulton, A., Mabey, J. E., Maudling, N., McGregor, L., Mitchell, A., Moulton, G., Paine, K., & Scordis, P. (2002). PRINTS and PRINTS-S shed light on protein ancestory. *Nucleic Acid Research,* **239-41** (1), 239–41.

Faulkner, D. & Jurka, J. (1988). Multiple aligned sequence editor (MASE). *Trends in Biochemical Science,* **13**, 321–2.

Galtier, N., Gouy, M., & Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequences alginemtn and molecular phylogeny. *Computer Applications for Biosciences,* **12**, 543–8.

Gille, C. & Frömmel, C. (2001). STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics,* **4**, 377–8.

Parrry-Smith, D. J., Payne, A. W., Michie, A. D., & Attwood, T. K. (1998). CINEMA - a novel Colour INteractive Editor for Multiple Alignments. *Gene,* **221** (1), GC57–63.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet,* **16** (6), 276–7.

Siepel, A., Farmer, A., Tolopko, A., Zhuang, M., Mendes, P., Beavis, W., & Sobral, B. (2001). ISYS: a decentralized component based approach to the integration of heterogeneous bioinformatics resources. *Bioinformatics,* **17** (1), 83–94.