Sequence analysis

Panta rhei (QAlign2): an open graphical environment for sequence analysis

Michael Sammeth, Thasso Griebel, Felix Tille and Jens Stoye*

Genome Informatics, Technical faculty, Bielefeld University, PO Box 10 01 31, 33501 Bielefeld, Germany

Received on December 4, 2005; revised on January 12, 2005; accepted on January 13, 2005
Advance Access publication January 17, 2006
Associate Editor: Dmitrij Frishman

ABSTRACT

Motivation: The first version of the graphical multiple sequence alignment environment QAlign was published in 2003. Heavy response from the molecular-biological user community clearly demonstrated the need for such a platform.

Results: Panta rhei extends QAlign by several features. Major redesigns on the user interface, for instance, allow users to flexibly compose views for multiple projects. The new sequence viewer handles datasets with arbitrarily many and arbitrarily large sequences that may still be edited by guided block moving. More distance-based algorithms are available to interactively reconstruct phylogenetic trees which can now also be zoomed and navigated graphically.

Availability: Executables and the JAVA source code are available under the Apache license at http://gi.cebitec.uni-bielefeld.de/qalign
Contact: qalign@cebitec.uni-bielefeld.de

1 INTRODUCTION

‘Panta rhei’ (everything flows), the famous words for which most historians give credit to Heraclitus (about 500 BC), carry the important message that all things change, while this change itself creates new things that otherwise could not exist. Like stated by this antique philosophy, often nature is not categorically decomposable into a set of static elementary components and sometimes many different aspects of the same data are to be gathered to puzzle together the mosaic of its meaning. Since we believe that also in biological sequence analysis a strong interaction between sequence alignment, phylogenetic analysis and visualization is required in order to produce meaningful results, we have chosen Panta rhei to be the name of the new version of QAlign.

Although for each of those tasks there already exist multiple free tools, they are still poorly associated with each other. On the one hand, there are state-of-the-art algorithms for multiple sequence alignment, which in the stand-alone version are available only as command-line tool, e.g. DCA (Stoye et al., 1997), DiAlign (Morgenstern et al., 1996) or T-COFFEE (Notredame et al., 2000). On the other hand, several tools exist to render multiple alignment layouts, e.g. SeaView (Galtier et al., 1996), CHROMA (Goodstad and Ponting, 2001) or JalView (Clamp et al., 2004). In most instances, these alignment viewers allow to scroll the layout and to apply coloring scheme decorations. However, most of them lack either editing capabilities, the ability to align sequences or to infer a tree. Also the recently developed graphical program MEGA3 (Kumar et al., 2004), that indeed provides a strong phylogenetic module, has only rather limited possibilities for sequence editing and automated alignment.

2 NEW FEATURES

In our opinion the integration of multiple algorithms for sequence alignment, interactive phylogenetic analysis and visual comparison possibilities is the reason why the first version of QAlign (Sammeth et al., 2003) has drawn large attention across all fields of molecular biology (Jaffe, 2003): up to now we have counted over 2000 registration requests. Panta rhei takes the idea one step further in the direction of a universal graphical environment for sequence analysis and extends the features of QAlign by several major components, described in the following paragraphs.

Workspace administration. In contrast to the previous version, it is now possible to work simultaneously on more than one sequence set; multiple projects may coexist, each of them based on specific input sequences. Within each project, different alignments may be created either manually, by an automated method or a combination of both. Phylogenetic trees and boxes derived from the alignment(s) are stored in folders of the corresponding project, may be navigated by browsing a hierarchical tree, and can be exported for further continuation (Fig. 1, left).

Visual composition of the user interface. The integration of multiple projects created the need for a more flexible user interface to view and compare the alignments and trees. While traditional programs for multiple sequence analysis often use multiple windows to display different modules, we have integrated a one-container layout, the flexdock libraries (http://flexdock.dev.java.net). These libraries allow to conveniently stock multiple containers in the same window that can be easily accessed or compared at any time. Docking points allow to compose every desired tiling of the screen (Fig. 1, background).

Multiple alignment algorithms. The spectrum of multiple alignment algorithms in QAlign is now extended by consistency-based heuristics like ClustalW (Thompson et al., 1994), DiAlign (Morgenstern et al., 1996) and T-Coffee (Notredame et al., 2000). For all the methods, parameters can be set interactively, and the progress of the algorithmic run can be monitored in the algorithmic console (Fig. 1, bottom-left).

Sequence editor. In view of the rapidly increasing number of genome projects, it becomes obvious why sequence analysis is performed increasingly on a semi-genomic or on a genomic scale. Hence, also modern sequence viewers should be capable...
of displaying genomic-scale data; the latter is not an easy task if editing functionality is to be provided. *Panta rhei* includes a sequence editor with a dynamical late-loading strategy that keeps track of editing events until they are written to disk (Fig. 1, center). By this, a nearly continuously fluid scrolling is guaranteed even for very long sequences. To test our visual component, we successfully loaded and modified the complete human chromosome 1 sequence (~300 Mb of DNA), whereas MEGA3 and SeaView were unable to process this dataset (Windows platform with 700 MB RAM).

*Tree inferring algorithms.* In addition to the Neighbor-Joining method, tree reconstruction algorithms for ultra-metric data have been added, i.e. the UPGMA, WPGMA and hierarchical clustering (Fig. 1, center-right).

*Phylogenetic tree viewer.* Figure 1 (right) depicts the new rendering component that allows to zoom the tree. A node inspector visualizes distance information, even if the corresponding subtree has been compacted. Like in QAlign, trees change automatically whenever the underlying alignment layout is modified; to enhance transparency, the sequences may also be re-ordered according to their occurrence in the tree. Furthermore, extended possibilities for the graphical export are offered via the freehep-libraries (http://java.freehep.org). To be specific, now additional pixel-based formats (e.g. BMP, PNG, GIF, JPG, etc.) and vector formats for drawing tools (e.g. EPS, EMF, SVG) are available.

### 3 CONCLUSION

*Panta rhei* is an extendable workbench for sequence analysis, and we invite other programmers to integrate their ideas in the graphical framework.

### ACKNOWLEDGEMENTS

The work was supported by a fellowship of the Ernst-Schering Research Foundation to M.S.

*Conflict of Interest:* none declared.

### REFERENCES


