QAlign: quality-based multiple alignments with dynamic phylogenetic analysis

M. Sammeth\textsuperscript{1,2,3,*,†}, J. Rothgänger\textsuperscript{4}, W. Esser\textsuperscript{1}, J. Albert\textsuperscript{1}, J. Stoye\textsuperscript{3} and D. Harmsen\textsuperscript{5}

\textsuperscript{1}Department of Computer Science II, University of Würzburg, 97074 Würzburg, Germany, \textsuperscript{2}International Graduate School of Bioinformatics, Bielefeld University, 33594 Bielefeld, Germany, \textsuperscript{3}Genome Informatics, Department of Technology, Bielefeld University, 33594 Bielefeld, Germany, \textsuperscript{4}RIDOM bioinformatics, 97082 Würzburg, Germany and \textsuperscript{5}Institute for Hygiene, University of Münster, 48149 Münster, Germany

Received on December 11, 2002; revised on February 12, 2003; accepted on March 4, 2003

\section*{ABSTRACT}
\textbf{Summary:} Integrating different alignment strategies, a layout editor and tools deriving phylogenetic trees in a ‘multiple alignment environment’ helps to investigate and enhance results of multiple sequence alignment by hand. QAlign combines algorithms for fast progressive and accurate simultaneous multiple alignment with a versatile editor and a dynamic phylogenetic analysis in a convenient graphical user interface.

\textbf{Availability:} QAlign is freely available over the internet at http://www.ridom.de/qalign/. The platform-independent JAVA technology used provides distributions for various operating systems and hardware architectures.

\textbf{Contact:} qalign@ridom.de

\section*{IMPLEMENTATION}
Due to its modular and layered structure, our program QAlign may easily be extended to support additional algorithms for both, multiple alignment and phylogenetic reconstruction. Herein we outline the features included in the current version.

\textbf{Multiple alignment algorithms}
QAlign is a new graphical environment integrating multiple features in the construction of the best multiple alignment for a specific set of sequences (FASTA and MSF sequence format supported). The algorithms monitor controls the construction of the multiple alignment where a fast progressive or a more accurate simultaneous approach may be chosen to align the sequences or parts of them (see Fig. 1 right). The heuristics used in the progressive approach (QAlign uses the variant of the MSA protocol, Gupta \textit{et al.}, 1995) of global multiple sequence alignment allow the alignment of even very large data sets. However, the drawback is that the resulting alignment is a fast approximation of the solution (McClure \textit{et al.}, 1994; Hickson \textit{et al.}, 2000).

In addition, QAlign contains an efficient and stable re-implementation of the NCBI’s MSA (multiple sequence alignment) program (Gupta \textit{et al.}, 1995). This is based on the simultaneous alignment strategy, an exact algorithm capable of finding the optimal mathematical solution. In addition to the optimizations used in MSA, the divide-and-conquer algorithm DCA (Tönges \textit{et al.}, 1996; Stoye \textit{et al.}, 1997) was used to achieve the simultaneous alignment of larger data sets. The desired quality-time tradeoff ratio for simultaneous alignment construction can be balanced by a slider. Both the progressive and the simultaneous alignment strategy may be used in a complementary manner on the same alignment layout.
Alignment editor features

After aligning the sequences, the graphical editor of QAlign provides features to analyse the result and modify the multiple alignment layout (as in Fig. 1, bottom). Gaps may be inserted or deleted and marked blocks may be moved within the alignment providing that the aligned sequences have the same length. An immediately updated consensus sequence with coloured bars shows the matching ratio of each column. These bars represent the conservation of different clusters across the alignment layout. They are also displayed as a bird’s eye view underneath a scrollbar thus allowing easy navigation to areas of low similarity. A secondary view is provided which may be used to extend the editor capabilities on one alignment or to compare two different multiple alignment layouts.

Dynamic phylogenetic analysis

A dynamic phylogenetic tree view makes visible the consequences of a change in the alignment with regard to the phylogenetic relationship (see Fig. 1, top) where branch lengths may change and nodes may swap according to the neighbour-joining method (Saitou and Nei, 1987). The tree may also be bootstrapped at any time to reveal its current stability. Thus, a phylogenetic reflection of the dynamics of the multiple alignment layout is obtained.

A variety of visual rearrangements is provided for the tree (e.g. subtrees may be collapsed or rearranged). Finally, the phylogenetic tree may be exported, either to a vectorial data format for drawing tools (SVG) or to a common format used by tree plotters (Newick).

CONCLUSION

QAlign provides a practical solution for the creation of refined multiple alignments: layouts produced by various algorithms may be used as a starting-point for changes done by hand, while the phylogenetic consequences are visualised on the fly. Furthermore, the comparison of multiple alignments is made easier because of the two alignment views integrated in the user interface of QAlign.

REFERENCES


