GeneAlign: Java Application Software for Multiple Sequence Alignment Editor

Ken Kurokawa [1]
ken@gen-info.osaka-u.ac.jp

Yoshitaka Hasegawa [1]
yoshi@gen-info.osaka-u.ac.jp

Kenichi Kawashima [2]
kawasima@hankyu.co.jp

Toshiyuki Okumura [2]
okumura@hankyu.co.jp

Tatsuya Takagi [3]
ttakagi@phs.osaka-u.ac.jp

Teruo Yasunaga [1]
yasunaga@gen-info.osaka-u.ac.jp

[1]   Genome Information Research Center, Osaka University, 3-1 Yamadaoka, Suita,
      Osaka 565-0871, Japan
[2]   Net Navi, 1-16-1 Shibata, Kita-ku, Osaka 530-0012, Japan
[3]   Graduate School of Pharmacy, Osaka University, 1-6 Yamadaoka, Suita,
      Osaka 565-0871, Japan

1   Introduction

Multiple sequence alignment is the basis for sequence analysis. If multiple sequence alignment is not correct, the results of the further analysis might be meaningless. Though we can easily perform multiple sequence alignment by several multiple sequence alignment programs, usually we have to modify the alignment in detail by hands for a further analysis.

For this purpose, we have developed GeneAlign which enable us to modify the sequence alignment mainly using mouse buttons in addition to automatic alignment using clustalw [1].

2   GeneAlign

GeneAlign is a multiple sequence alignment editor written in java language and has excellent graphical user interface (Fig. 1). GeneAlign can treat both nucleotide sequences and amino acid sequences. When we want to perform alignment of coding regions of nucleotide sequences, we can choose one of the two modes, nucleotide sequence mode or coding region mode. In the former mode, GeneAlign performs multiple sequence alignment of nucleotide sequences. In the latter mode, coding regions are translated into the corresponding amino acid sequences and the regions are aligned as amino acid sequences, though the resultant alignment is given in nucleotide sequences.

GeneAlign has the following features that support handling multiple sequence alignment.

1. Mouse left button click on a base site or an amino acid site inserts a deletion and shift the right part of the sequence from the mouse cursor to the right. Mouse left button click with pressing shift key inserts a deletion and shift the left part of the sequence from the mouse cursor to the left.

2. Mouse right button click on a deletion site deletes it. Nothing occurs clicking the button on a base site or an amino acid site.

3. In the above operations, we can select more than two sequences with mouse dragging and can insert or delete deletions to the selected sequences at once.

4. We can lock the correct parts of alignment. Any operation can not be performed in the lock parts. An unlocked part between adjacent lock parts can be performed any operations including automatic alignment using clustalw.
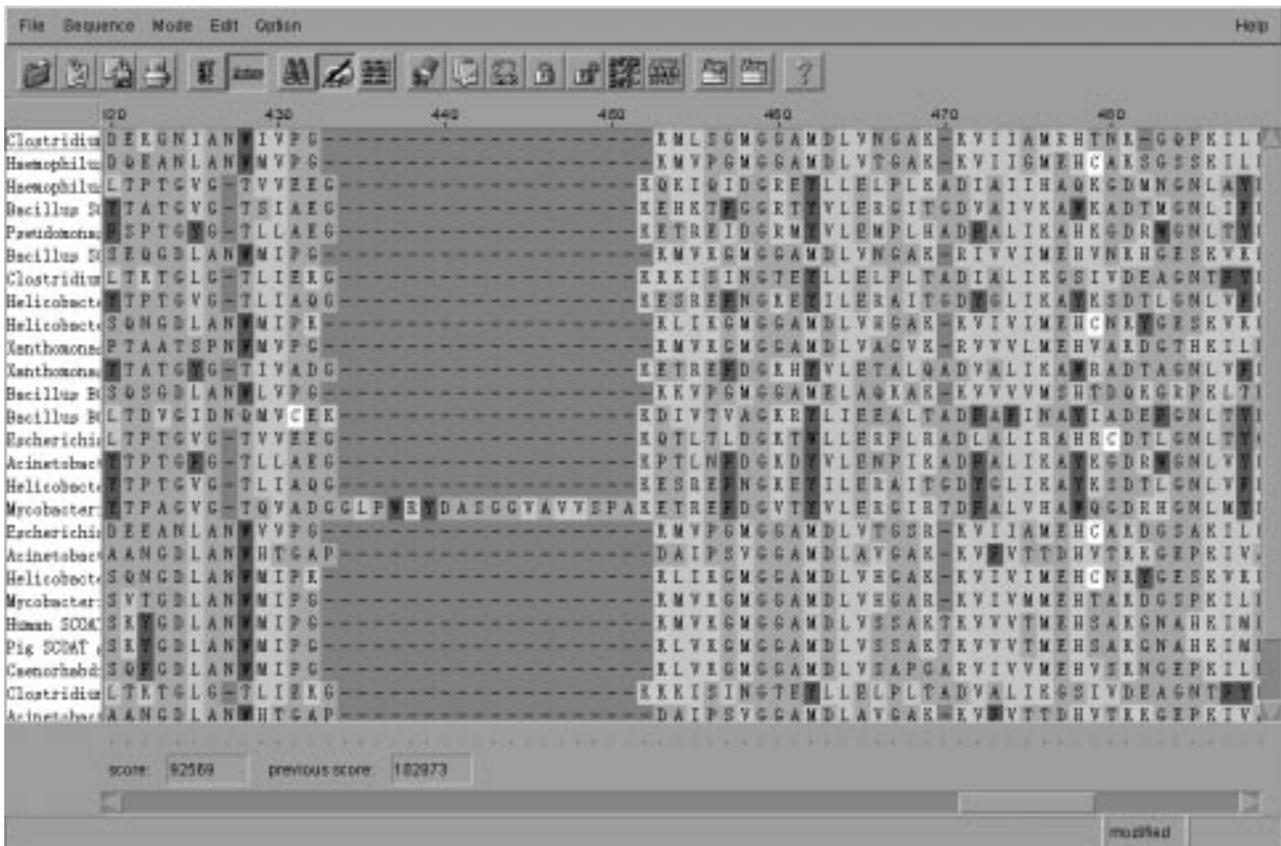
Figure 1: An example of multiple sequence alignment with GeneAlign.

5. Clicking Undo icon, the previous operation is cancelled. Furthermore continuing clicking Undo icon, we can go back to any stage of operations.

We have developed GeneAlign as Java application mainly because Java application software is able to read/write local files and to run on many platforms such as Unix workstations, Windows PC and Macintosh. Though we were worried about its speed for operations, GeneAlign can perform any operation at comfortable speed. GeneAlign can be downloaded through our Web page (http://www.gen-info.osaka-u.ac.jp/geneweb2/genealign/) and the instruction for installation and the user manual are also available on the same page.

References

[1] Thompson, J.D., Higgins, D.G. and Gibson, T.J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, *Nucleic Acids Research*, 22:4673–4680, 1994.