Multiple alignments programs
Common tools

The EMBOSS package: kalign, ClustalW, Muscle, MAFFT, T-Coffee

http://www.ebi.ac.uk/services/dna-rna
**Common idea: Progressive alignment**

**The essential idea:** Add the sequences one at a time producing intermediate alignments that are subsequently used to build the final multiple alignment ...

---

**The general method**

Compute all pairwise alignments. This gives all pairwise distances.

Build “guide tree”, e.g. using the NJ-methods

Construct multiple alignment by aligning sequences and alignments in the order described by this tree.
Progressive alignment: Clustal W

Idea: Build a multiple alignment \( M \) for a family \( \mathcal{F} = \{ S_1, \ldots, S_n \} \) by merging alignments of subsets of \( \mathcal{F} \).

Questions:
1) Which subsets to merge?
2) How to merge two alignments?
Answers:

1) Build a "guide tree" reflecting the evolution tree of the sequences.

At each internal node, build an alignment for $F_1 \cup F_2$ by merging these alignments...

2) Merge alignments by aligning them as we align sequences. The alignments are considered as "sequences"...
Clustal W: Algorithm

Step 1 - Build the guide tree

* Compute distance matrix $D_{ij} = D(S_i, S_j)$,

where $D(S_i, S_j)$ is the optimal score of an alignment of $S_i$ and $S_j$

Time: $O(k^2 \cdot n^3)$

* Build Neighbor-joining tree for $F$ using $D_{ij}$...

Time: $O(k^3)$
Clustal W: Algorithm

Step 2 - Build alignment by merging up through \( T \)

We merge two alignments by aligning them as "sequences" using global alignment methods...
Clustal W: Algorithm

Step 2 - Build alignment by merging up through $T$

We merge two alignments by aligning them as "sequences" using global alignment methods...

\[
\begin{array}{c}
\text{AC} \\
\text{AC} \\
\end{array}
\quad \begin{array}{c}
\text{TT} \\
\text{TT} \\
\end{array}
\quad \Rightarrow
\begin{array}{c}
\text{ACAC} \\
\text{ACAC} \\
\end{array}
\]

Cost of induced pairs:
\[
\text{cost} \left( \begin{array}{c}
\text{AC} \\
\text{AC} \\
\end{array} \right) = s(A,A) + s(A,C) + s(A,-) + s(C,A) + s(C,C) + s(C,-)
\]
Clustal W: Algorithm

**Step 2 - Build alignment by merging up through**

* We merge two alignments by aligning them as "sequences" using global alignment methods...

\[
\begin{align*}
\begin{bmatrix}
A & A & T & T \\
A & C & C & G \\
A & - & - & C
\end{bmatrix},
\begin{bmatrix}
A & A & T & T \\
A & C & C & G \\
A & - & - & C
\end{bmatrix}
\Rightarrow
\begin{bmatrix}
A & A & - & - & T & T \\
A & C & C & G & T & T \\
A & A & C & - & - & T \\
\end{bmatrix}
\end{align*}
\]

Cost given by gapcost function, e.g. \( g(k) = \alpha + \beta \cdot k \).
Clustal W: Algorithm

Running time? Depends on the length of the alignments produced up through the guide tree. Best case $O(kn^2k^2)$, worst case $O(k(nk)^2k^2)$