16 Multiple Match Refinement and T-Co ee

In this section we describe how to extend the match refinement to the multiple case and then use *T-Co ee* to heuristically compute a multiple trace.

This exposition is based on the following sources, which are all recommended reading:

1. Notredame, Higgins, Heringa: T-Coffee, a Novel Method for Fast and Accurate Multiple Sequence Alignment, Journal of Molecular Biology, 2000, Vol 302, pages 205-217.

2. Rausch, Emde, Weese, Döring, Notredame, Reinert: Segment-based multiple sequence alignment, ECCB 2008, Cagliari

16 Multiple Match refinement

We start by extending the pairwise match refinement to the multiple case. The result of the multple match refinement naturally induces an input graph for the *multiple trace problem* which is NP-hard but could be solved with ILP based techniques.

However, in practice we can resort to more efficient, but heuristic methods. In the second part of the lecture we will describe the TCoffee algorithm which is basically a heuristic for the multiple trace problem (although originally not advertised as such).

16.1   Multiple Match refinement

Lets start by extending the definitons from the pairwise case to the multiple case.

Definition 1. Let $\mathcal{S} = \{S^1, S^2, ..., S^n\}$ be a set of n sequences with $S^p = s^p_1 s^p_2, ..., s^p_{|S^p|}$, $p \in \{1, 2, ..., n\}$. A *segment match* $M = (S^p_{ij}, S^q_{kl})$ is an alignment between the two *segments* $S^p_{ij} = s^p_{i+1} s^p_{i+2} ... s^p_j$ and $S^q_{kl} = s^q_{k+1} s^q_{k+2} ... s^q_l$ with $p, q \in \{1, 2, ..., n\}$ and $p \quad q$ and $0 \le i \le j \le |S^p|$ and $0 \le k \le l \le |S^q|$.

The only change is that we consider more than two sequences. The other definitions are the same except that we consider now matches between all pairs of sequences. We refrain from giving them here again.

As in the pairwise case, the *segment match refinement* algorithm takes as input a set of segment matches $\mathcal{M}$ containing pairwise segment matches for a set of sequences $\mathcal{S} = \{S^1, S^2, ..., S^n\}$ together with a set of projection maps   . The output is the *minimal resolved refinement* $\mathcal{M}'$ of $\mathcal{M}$.

For example the four depicted segment matches in the next figure (a) would be refined into a total of seven segment matches in the figure (b). The figure already shows the corresponding input alignment graph for the trace problem. One can clearly see the reduction in problem size compared to the non segment match alignment graph version.