

Invited Talk

Combining Sequence Information with T-Coffee

Cedric Notredame
CNRS, Marseille, France

Well integrated biological data lends itself to the identification of biologically meaningful patterns. Multiple Sequence Alignments constitute one of the most powerful ways of carrying out such a task. In this context, the integration takes the form of simultaneously aligning related sequences in order to reveal evolutionary conserved patterns. Multiple Sequence Alignments have so many applications that they have become household items in biology and few data processing pipelines exist that do not require the assembly of an alignment. Yet, the wealth of available alternative methods means that the user is not only faced with the problem of selecting and aligning sequences, but also with the necessity of choosing one method or integrating the results delivered by many. In the course of this seminar I will discuss how various methods can be integrated into one. I will also go further and show that a multiple sequence alignments can be used to integrate much more than sequence information, as long as this information is properly mapped onto the sequences. This concept, named template-based multiple sequence alignment will be illustrated with a simple example: the combination of sequences and structures within multiple sequence alignments. I will finally discuss how multiple sequence alignment methods are currently validated and why I believe we need to challenge these procedures in order to take further our understanding of biological sequences. Most of the tools discussed in this talk are available from www.tcoffee.org.

References:

1. Armougom, F., S. Moreti, O. Poirot, S. Audic, V. Kedaous, and C. Notredame, *Expresso: Automatic Combination of Sequence and Structural Information*. Nucleic Acids Res, 2006. **In Press**.
2. Wallace, I.M., O. O'Sullivan, D.G. Higgins, and C. Notredame, *M-Coffee: combining multiple sequence alignment methods with T-Coffee*. Nucleic Acids Res, 2006. **34**(6): p. 1692-9.
3. O'Sullivan, O., K. Suhre, C. Abergel, D.G. Higgins, and C. Notredame, *3DCoffee: combining protein sequences and structures within multiple sequence alignments*. J Mol Biol, 2004. **340**(2): p. 385-95.
4. Notredame, C., D.G. Higgins, and J. Heringa, *T-Coffee: A novel method for fast and accurate multiple sequence alignment*. J Mol Biol, 2000. **302**(1): p. 205-17.