Molecular Evolution and Selection Analysis

The central hypothesis of comparative bioinformatics is that a structure that some sequences posses is more conserved than the sequences themselves. Since different structural elements have different evolutionary pattern, we can predict the joint structure of a set sequences by analyzing their evolution when the central hypothesis is true.

The background philosophy behind the central hypothesis of comparative bioinformatics is the Kimura's neutral evolution theorem. Indeed, in comparative structure prediction, we assume that a substring possessing a structural element undergoes a neutral evolution typical for the secondary structure element. For example, it is well-known that substitutions are more frequent in loop regions of a protein than in beta-sheets. Insertion-deletions typically go by three in coding regions while they might have an arbitrary length in intronic sequences, etc.

However, the structure might also change during the evolution. When the structure is being changed, the sequence usually undergoes positive selection. The easiest way to detect positive selection is to measure the ratio of synonymous and non-synonymous mutations in coding regions. A synonymous mutation is supposed to be a neutral mutation, since it does not change the protein sequence itself. (However, it changes codon frequency hence, speed of translation, so might change the co-translation folding, moreover, it might change local RNA secondary structure in mRNA, see Meyer & Miklos (2005) and citations in it.) The ratio of synonymous and non-synonymous mutations tells us if there is positive selection in the selected sequences.

Exercises

1. Read the paper: Laurence D. Hurst: The Ka/Ks ratio: diagnosing the form of sequence evolution
2. Download and install PAML: http://abacus.gene.ucl.ac.uk/software/paml.html#download
3. Analyze the CHRNA5, CHRNA3, and CHRNB4 genes with PAML for positive selection. See the PAML manual for further instructions: http://abacus.gene.ucl.ac.uk/software/pamlDOC.pdf

References

- Meyer, I.M. & Miklós, I. (2005) Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. Nucleic Acids Research, 33(19)6338-6348.
- Laurence D. Hurst (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution TRENDS in Genetics Vol.18 No.9 486-487.