Molecular Phylogenetics and Evolutionary analysis

Tutorial proposal for ISMB2003

Tutor

Graziano Pesole is an Associate Professor at the University of Milan (Italy). He has for many years performed nucleotide and amino-acid sequence data analysis aimed at the recognition of candidate patterns with functional roles. His research interests also include molecular evolution. In particular, his interest in functional analyses of untranslated regions of eukaryotic mRNAs has lead to the development of a specialised database. He has developed analysis software and several algorithms widely used by the scientific community and available also through web browsers. He has also been the coordinator of European Union funded projects and has filed an international patent for the selection of primers for RNA fingerprinting. In the field of molecular evolution, he has contributed to the development of the Markov model (also known as General Time Reversible) and has carried out several studies on the evolution of the mitochondrial genome at both the intra and inter-species level, to clarify some aspects of the origin of modern man, and to reconstruct mammal phylogeny.

He is the author of over 80 papers published in international journals and co-author of books on Bioinformatics and Genomics published (or in press) by italian (Zanichelli, Gnocchi) and international (Wiley) editors.

He is associate editor for the following international journals: GENE, Briefings in Bioinformatics, Journal of Molecular Evolution.

The proposed tutorial draws from experience gained during seminars, graduate and PhD courses in the field of Molecular Evolution.

Length

Half day

Tutorial presentation

The purpose of this tutorial is to provide the student with an introductory survey of the field of molecular evolution, phylogenetics and molecular taxonomy.

Molecular evolution represents the study of informational macromolecules (DNA and proteins) in a temporal framework and allows the reconstruction of the evolutionary history of genes and organisms. Understanding the evolutionary history of a gene or a gene family is crucial to the *in-silico* inference of its function and the prediction of the effects of its inactivation. Indeed, the study of evolutionary dynamics and mechanisms certainly represents a valuable contribution to the understanding of the structure-function relationships of genetic material. The sequencing of complete genomes from prokaryotes, eukaryotes and organelles (mitochondria and chloroplasts) has offered new perspectives that may ultimately allow us to unravel the mechanisms underlying the evolution of the genome as a whole and to better understand the relationships between genotype and phenotype.

This tutorial will provide a survey of mechanisms driving molecular evolution at the level of genes, their products and of genomes as well describing basic methodological tools for comparative and phylogenetic analyses of molecular data. The goal is to provide students with a basic expertise in evolutionary and phylogenetic analysis and most importantly, to give them the ability to see the evolutionary relevance of molecular data.

Tutorial Outline

1. Introduction (slide 1-3)

2. Molecular mechanisms underlying molecular evolution: point mutations, deletions and insertions, inversions and other kind of rearrangements; mutation and fixation (slides 4-17).

3. Homology, Orthology and Paralogy (slides 18-20)

4. Use of DNA vs protein sequences in evolutionary analyses (slides 21-23)

5. Multiple alignment is the critical input data for evolutionary analysis (slides 24-25)

6. Estimating genetic distances between homologous sequences (slides 26-43)

    6.1 Introduction to stochastic models (slide 28)

    6.2 Stationarity in base composition (slides 29-30)

    6.3 Substitution models: Jukes-Cantor's one parameter model (slides 31-33)

    6.4 The Kimura two-parameter model (slide 34)

    6.5 The General Time Reversible model (slide 35)

    6.6 Outline of substitution models (slide 36)

    6.7 Accounting for invariant sites and site rate heterogeneity (slides 37-39)

    6.8 Which model is best to use? (slide 40)

    6.9 How to estimate model parameters? (slide 41)

    6.10 Distance measures for protein sequences (slide 42)

    6.11 Substitution/site is the measure unit for genetic distance (slide 43)

7. Estimating site-specific rate variability for nucleotide and protein sequences (slides 44-45)

8. Inferring Molecular Phylogeny

    8.1 Introduction to trees (slides 47-48)

    8.2 Rooted and unrooted trees (slide 49)

    8.3 Cladograms and Phylograms; Newick format to represent trees (slides 50-51)

    8.4 Classification of tree-building methods (based on data and method) (slide 52)

    8.5 Distance methods: UPGMA (slide 53)

    8.6 The Molecular Clock (slides 54-56)

    8.7 Neighbor-joining (slide 57)

    8.8 Drawbacks of clustering methods (slide 58)

    8.9 Maximum parsimony (slide 59-60)

    8.10 Maximum likelihood (slides 61-62)

    8.11 Testing alternative hypotheses using ML (slide 63)

    8.12 Assessing tree reliability: Bootstrap and Jacknife (slides 64-65)

9. Softwares and packages for Molecular Phylogeny (slides 66-67)

Intended Audience

Computer scientists, biologists, statisticians with interest in methods and applications of evolutionary analysis of biological sequences (DNA and proteins). The tutorial aims to be introductory and requires only basic knowledge on probability, statistics and molecular biology.