



Università degli Studi di Verona

Scuola di Dottorato di Scienze Ingegneria e Medicina

Ca' Vignal 2  
Strada le Grazie 15  
37134 Verona - Italia  
Tel. +39 045 802 7069  
Fax +39 045 802 7068

**Dottorato di ricerca in Informatica**

**Graduate course in**

**Advanced algorithms for the analysis and visualization of  
DNA and protein sequences**

**July 16-26, 2007 (20 hours)**

**Zu-Guo Yu**

School of Mathematical Science  
Queensland University of Technology, Australia

**Week 1 – July 16/19, 2007**

**16/07 (14.30-17.30) - Biological sequence databases (3 hrs)**

- 1.1 Course introduction
- 1.2 Introduction to biological databases
- 1.3 Primary and secondary databases
- 1.4 Nucleotide sequence databases
- 1.5 Protein sequence databases
- 1.6 Protein structure databases
- 1.7 Genomic databases

**18/07 (10.30-12.30 and 14.30-16.30) - Predictive methods using DNA sequences and protein sequences and structure (4 hrs)**

- 2.1 Introduction
- 2.2 Gene prediction methods
- 2.3 DNA promoter analysis characterization and prediction
- 2.4 Predicting features of individual residues
- 2.5 Predicting function
- 2.6 Protein structure prediction
- 2.7 Protein structure evaluation
- 2.8 Protein structure comparison
- 2.9 Protein secondary structure prediction

**19/07 (14.30-17.30) - Pairwise and multiple sequence alignment (3 hrs)**

- 3.1 Introduction
- 3.2 Global versus local sequence alignments
- 3.3 Dot-plots
- 3.4 Score matrices
- 3.5 Multiple alignment
- 3.6 Structural and evolutionary alignment
- 3.7 How to multiply align sequences
- 3.8 How to generate multiple structural alignments

## Week 2 – July 23/26, 2007

### 23/07 (14.30-17.30) - Phylogenetic analysis without sequence alignment (3 hrs)

- 4.1 Introduction
- 4.2 Why learn more about phylogenetics and other evolutionary analyses?
- 4.3 Interpretation
- 4.4 How to construct a tree
- 4.5 SVD method
- 4.6 Markov model
- 4.7 Dynamical language model
- 4.8 Fourier method and log-correlation distance
- 4.9 Entropy density profile with log-correlation distance
- 4.10 Mutual information distance

### 25/07 (10.30-12.30 and 14.30-16.30) - Visualization of biological sequences (4 hrs)

- 5.1 Introduction
- 5.2 DNA walk model
  - 5.2.1 One-dimensional DNA walk
  - 5.2.2 Two-dimensional DNA walk
  - 5.2.3 Higher-dimensional DNA walk
- 5.3 Chaos game representation of biological sequences
  - 5.3.1 Fundamentals of chaos game theory
  - 5.3.2 Chaos game representation of DNA sequences
  - 5.3.3 Chaos game representation of protein sequences
  - 5.3.4 Chaos game representation of protein structures
  - 5.3.5 Chaos game representation of amino acid sequences based on detailed HP model
- 5.4 Two-dimensional portrait representation of DNA sequences
  - 5.4.1 Graphical representation of counters
  - 5.4.2 Fractal dimension of the fractal set for a given tag

### 26/07 (14.30-17.30) - Measures of biological sequences (3 hrs)

- 6.1 One-dimensional measure representation of biological sequences
  - 6.1.1 Measure representation of complete genomes
  - 6.1.2 Measure representation of linked protein sequences
  - 6.1.3 Measure representation of protein sequences based on detailed HP model
- 6.2 Z curve representation of DNA sequences
- 6.3 Number sequence of DNA sequences
- 6.4 Fractal method and RQA for protein structure classification

---

Lectures will take place at the  
**Sala Verde**  
**Dipartimento di Informatica**  
**Ca' Vignal 2 - Strada le Grazie 15, 37134 Verona, Italy**

The course is supported by the Università degli Studi di Verona, with a CooperInt grant and by the Dipartimento di Informatica.

The admission to the course is FREE. Attendees must submit an application via email to the Local organizer indicating their affiliation and current position by **July 9, 2007**.

Proficiency certificates or equivalent credits will be provided on request and upon completion of independent homework.

Local organization and contact: Federico Fontana  
federico.fontana@univr.it

Web info: [www.di.univr.it](http://www.di.univr.it)