

The ICGEB EMBnet node: AGM 2010 report



Sándor Pongor

ICGEB, Trieste, Italy

ICGEB, the International Centre for Genetic Engineering and Biotechnology (<http://www.icgeb.org>) is a non-profit international organization that provides scientific research and an educational environment for the benefit of developing countries. With laboratories in Trieste, Italy (headquarters), New Delhi, India and Cape Town, South Africa, the ICGEB forms an interactive network with Affiliated Centres in 39 of its 60 Member States. ICGEB is part of the United Nations System. At present, more than 400 people from 38 different countries are working in the ICGEB laboratories as research scientists, postdoctoral fellows, PhD students, research technicians and administrative personnel. Research and training programs are focusing on biomedicine, crop improvement, environmental protection/remediation, biopharmaceuticals and biopesticide production. External services include biosafety counselling, biotechnology transfer and bioinformatics.

Bioinformatics services

Bioinformatics was the first service started at ICGEB in 1990, established with the aim of providing access to sequence databases, computing facilities (GCG). Gradually, the emphasis has shifted; today ICGEBnet mainly offers courses and consultation in bioinformatics, as well as open bioinformatics services - these are mostly based on web servers developed within the Protein Structure and Bioinformatics Group which is responsible for the project. The use of ICGEBnet is available free of charge to all ICGEB Signatory Country scientists. Protein Tools include protein domain prediction from sequence and structure, as well as access to the SBASE protein domain library, a Benchmark collection for protein clas-

sification, and various services for the analysis of protein 3D structures. DNA Tools include servers to predict and visualize bent regions in DNA sequences, as well as plots of various structural parameters along the sequence. With a total of over 1300 students since 1990, the annual course "Bioinformatics: Computer Methods in Molecular Biology" offers lectures and practical sessions on biological sequence analysis as well as an introduction to the major bioinformatics services available worldwide. Participation is free for selected applicants from ICGEB Member States. ICGEBnet provides tutorials and consultation for the students of the Ph.D. course at ICGEB-Trieste.

At the moment, the main platform for our bioinformatics services is a Sun Fire V20z system, with 22 computing cores equipped with 8GB RAM. During the last few years, 1-2 people have been actively maintaining and developing the bioinformatics services at ICGEB. At present, there are 327 registered users.

Software development

Bioinformatics tools developed by the group:

DNA tools	
Bend.it ¹	This server predicts DNA curvature from DNA sequences.
Plot.it ²	This server plots various physicochemical, statistical or locally computed parameters along DNA sequences (1-D or sequence-plots) or against each other (2-D plots).
Model.it ³	This server produces a 3D model of a DNA molecule, given a sequence of maximum 700 nucleotides.
Introns ⁴	Intron phase pattern conservation server.

1 http://hydra.icgeb.trieste.it/dna/bend_it.html

2 http://hydra.icgeb.trieste.it/dna/plot_it.html

3 http://hydra.icgeb.trieste.it/dna/model_it.html

4 <http://hydra.icgeb.trieste.it/~kajla/introns/>

<u>Protein tools</u>	
Sbase ¹	Support vector machines, domain prediction system
Fthom ²	Predict domains in your sequence
P450 ³	A directory of p450-containing systems
CX ⁴	This server calculates the atomic Protrusion Index from a three-dimensional protein structure.
Pride ⁵	This server calculates the PProbability of IDentity between three-dimensional domains (or whole structures)
DPX ⁶	This server calculates the atom depth from a three-dimensional protein structure.
Benchmark ⁷	A protein classification benchmark collection for testing machine-learning algorithms
prideNMR ⁸	NMR Protein fold similarity server
theGPM ²	This is an in-house modified version of the Gpm global proteome machine, a tandem mass spectrometry data analysis server.

1 <http://hydra.icgeb.trieste.it/sbase/>

2 <http://hydra.icgeb.trieste.it/sbase/sbase.php?sec=analyse&sub=predict>

3 <http://www.icgeb.org/~p450srv/>

4 <http://hydra.icgeb.trieste.it/cx/>

5 <http://hydra.icgeb.trieste.it/pride/>

6 <http://hydra.icgeb.trieste.it/dpx/>

7 <http://net.icgeb.org/benchmark/>

8 <http://net.icgeb.org/pridenmr/>

9 http://proteome.icgeb.trieste.it/tandem/thegpm_tandem.html