

Training Mexican scientists and students on MRS/EMBOSS: A course report



George Vasilios Magklaras¹, Romualdo Zayas²

¹The Biotechnology Centre of Oslo, The University of Oslo, Oslo, ²UNAM's Bioinformatics National Node, Center for Genomic Sciences, Cuernavaca, Mexico

Between March 22nd and March 26th 2010, the National Bioinformatics Node of Mexico (celebrating its 10th anniversary) at the Center for Genomics Sciences/UNAM, in association with the Norwegian EMBnet node at the University of Oslo/Norway, gave a course entitled "Sequence Mining in Sequence Databases: A case with MRS and EMBOSS". The Norwegian EMBnet node provided the instruction material and access to their services, and the Mexican EMBnet node the workstation laboratory and access to course hand-out material. This was one of the first MRS [1] courses given outside European borders, focusing on teaching the course participants a range of basic concepts about various bioinformatics databases, as well as how to access them in a programmatic way, in order to construct useful sequence processing pipelines for their research.

The course ran as a full day (9:30 – 15:00 hours) activity and it was very much a "hands-on" laboratory. Each day consisted of 1.5/2 hour morning lecture followed by a small break and a carefully prepared practical tutorial session, where students could practice solving real-world cases in the lab and ask questions about the tools. The first day was devoted to the basics of sequence databases and formats, as well as pipeline construction paradigms (comparing the command line approach with other approaches, such as Taverna [2], Accelrys Pipeline Pilot[3]). The next topic was a gentle introduction to key EMBOSS [4] applications and attempts to access specific



Figure 1. Group picture of course attendants

sequence databases. The remaining days were dedicated to introducing MRS as a sequence retrieval tool and the task of combining EMBOSS and MRS applications to construct complex pipelines. During the final day, the students had access to the laboratory, in an attempt to construct pipelines useful for their research problems under the guidance of the course instructor.

In summary, the course achieved its objectives and gave students a useful experience in using both EMBOSS and MRS in combination with scripting languages to construct their own pipelines. This represents a nice example of the EMBnet community at work.

In addition, aside from course instruction, the EMBnet Technical Management Project Committee is in the process of helping the Mexican node to install most of the tools (EMBOSS/MRS 4) in their computational infrastructure, so by the time you read these lines, Mexico will have a working MRS 4 installation serving their community.



Figure 2. "Hands on" laboratory in the undergraduate program classroom of the Center for Genomics Sciences

Acknowledgments

The node managers would like to thank the following people for their support in the course:

- Harald Dahle, for setting up the student accounts and the environment on the EMBnet Norway side.
- Alfredo Hernández Alvarez and the technicians of the LCG UNAM IT unit for recording the course and making the laboratory equipment run smoothly for the entire duration of the course.

References

1. Hekkelman M L and Vriend G. (2005) MRS: a fast and compact retrieval system for biological data. *Nucleic Acids Research* 33 (Web Server issue):W766-W769
2. Hull D, et al. (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Research* 34 (Web Server issue):W729-W732
3. Accelrys, Inc., 10188 Telesis Court, Suite 100, San Diego, CA. URL: <http://accelrys.com/products/pipeline-pilot/>
4. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16: 276-277