Portuguese EMBnet node: AGM2011 report

Pedro L Fernandes

Instituto Gulbenkian de Ciência, Oeiras, Portugal

The activity of the Portuguese node of EMBnet is split in two areas: service provision and training.

Service Provision

A re-equipment grant (REEQ/741/BIO/2005 from Fundação para a Ciência e Tecnologia, the national funding agency for science) has allowed the Instituto Gulbenkian de Ciência to move in the direction of High Performance Computing (HPC). The Portuguese node of EMBnet migrated its users into the new computing platform, a 60-node Linux cluster based on IBM Blade Centers in March 2006. Satisfying the needs of about 230 users encompassed two spheres of action: one for users who were not involved in HPC itself but wanted to remain operational in their day-to-day business; and another for 'power users' who wished to implement solutions using the parallel architecture features (via MPI). We continued to provide full support for EMBOSS and other non parallelised code via the Inquiry Portal, and everybody gained availability as the use of non-concurrent Linux machines came to the rescue. The portal offers several flavours of BLAST that use the MPI and run on up to 8 blades in parallel, and integrates locally developed software seamlessly, to the point of automatically generating local web services. The cluster is used to support long-term simulations via ssh, which is mostly used for Population Genetics simulations. Management of the cluster is automated by scripts and by monitoring tasks that require very little attention [Fernandes, 2006]. The cluster usage requires minimal maintenance in terms of replacement of parts, which is eased by the redundancy of the hardware and electrical power.

The installation is named Centro Português de Bioinformática – Recursos de Alta Prestação.

Figure 1. A practical, "Proteomics Data Analysis" course in 2010.

Training

The IGC has a permanent Bioinformatics Training Programme (GTPB) that has been running regularly since 1999. This Programme [Fernandes, 2010] presently holds 16 hands-on training courses per year. The attendance is international, and though the travel costs make it more affordable for local trainees, it is more widely accessible, as local lodging is also available at very low rates. Care has been taken to generate expansion plans and testing solutions for handling remote audiences, b-learning, and lecture capture. Future plans include testing telepresence solutions to better approximate face-to-face situations. The GTPB offers an Introductory course once or twice per year, several thematic courses and "Foundations" type courses, aimed at recycling fundamental knowledge like Statistics, Database Design, *etc.* We also offer a "Bioinformatics using Python for Biologists" course, which fosters the development of programming skills using a thematic set of bioinformatics problems. Approximately 300 people now benefit from this programme annually, bringing the total to ~2,000 trainees in the last 12 years.

The activities of the Portuguese node contribute to the integration of users, and communities of users, in global communities and activities; EMBnet-originated software, Quick Guides, newsletter articles, etc. are frequently used. The educational experiments that are under way in GTPB are shared via the planned activity of the ET PC, and it is foreseeable that the resulting actions can benefit the whole EMBnet community.

Pedro Fernandes, the node manager, has served as chair in the P&PR PC (1999-2010) and as chair of the ET PC since June 2010. He is a founding member of the BTN, a thematic group that studies Bioinformatics Training techniques [Schneider *et al.*, 2010].

Figure 2. The new Bioinformatics Training Room at the Instituto Gulbenkian de Ciência.

Publications

Fernandes, P.L. (2006) Single handed node management. EMBnet.news 12, 19-20.

Fernandes, P.L. (2010) The GTPB training programme in Portugal. Brief. Bioinform. 11, 626-634, doi:10.1093/bib/bbq063.

Schneider, M.V., Watson, J., Attwood, T.K., Rother, K., Budd, A., McDowall, J., Via, A., Fernandes, P., Nyronen, T., Blicher, T., Jones, P., Blatter, M-C., De Las Rivas, J., Judge, D.P., van der Gool, W. and Brooksbank, C. (2010) Bioinformatics training: a review of challenges, actions and support requirements. Brief. Bioinform. 11, 544-551, doi:10.1093/bib/bbq021.

Websites
- http://hermes.igc.gulbenkian.pt
- http://gtpb.igc.gulbenkian.pt

Annex

List of GTPB courses in 2010:
- Molecular Evolution, Phylogenetics and Adaptation
- Introductory Bioinformatics (First Course)
- Bioinformatics using Python for Biologists
- Hunting for genes and promoters

- Microarray Data Analysis using GEPAS and Babelomics
- Automatic Functional Annotation and Data Mining
- Macromolecular NMR assignment with CcpNmr Analysis
- Analysis and manipulation of phylogenomic data using ETE
- RNA Bioinformatics
- Microarray Data Analysis using R and Bioconductor
- Proteomics Data Analysis
- Structural Genomics and Drug Design
- Pathway Analysis and Drug Targets
- Integrative Cancer Genomics
- Biostatistical Foundation in Bioinformatics
- Introductory Bioinformatics (Second Course)

List of GTPB courses in 2011:
- Genetic Architecture of Complex Traits
- Massive Data Analysis (using Babelomics)
- Automatic Functional Annotation and Data Mining
- Introductory Bioinformatics
- Model-based Inference of Population Structure
- Molecular Evolution, Phylogenetics and Adaptation
- Bioinformatics using Python for Biologists
- RNA Bioinformatics NGS Data Management
- Hunting for Genes and Promoters
- Building Bioinformatics Web Applications
- Knowledge Discovery and Management in Chemoinformatics
- Microarray Data Analysis using R and Bioconductor
- Integrative Approaches in Genomics
- Biostatistical Foundations in Bioinformatics