Book Title
Book Editors
IOS Press, 2003

Biomedical Aplications in EELA[1]

Miguel Cardenas [a], Vicente Hernández [b], Rafael Mayo [b], Ignacio Blanquer [b], Javier Perez-Griffo [a], Raul Isea [c], Luis Nuñez [c], Henry Ricardo Mora [d] and Manuel Fernández [d]

[a] *Extremadura Advanced Research Center (CIEMAT)*
[b] *Universidad Politécnica de Valencia, ITACA - GRyCAP*
[c] *Universidad de los Andes*
[d] *Cubaenergía*

Abstract. The current demand for Grid Infrastructures to bring collabarating groups between Latina America and Europe has created the EELA proyect. This e-infrastructure is used by Biomedical groups in Latina America and Europe for the studies of ocnological analisis, neglected diseases, sequence alignments and computation plygonetics.

1. Introduction

Funded by the European Commission, the EELA Project ("E-Infrastructure shared between Europe and Latin America") builds a digital bridge between the existing e-Infrastructure initiatives that are in process of consolidation in Europe (in the framework of the European EGEE Project), and those that are emerging in Latin America, throughout the creation of a collaborative network that share an interoperable Grid infrastructure to support the development and test of advanced applications.

EELA aims to position the Latin American countries at the same level of the European developments in terms of e-Infrastructures. Now that the network infrastructure in Latin America is stable, the EELA focus will be in the Grid infrastructure and in some related e-Science applications. Therefore, the project's participant institutions have identified two fundamental scopes: the creation of a human network in e-Science - valuing its necessities and giving training to it - and the conduction of the technological developments that will allow Grid development and operation in the region. Establishing a collaborative network of research scientist and people

This document describes the biomedicine applications which are currently deployed on the pilot EELA infrastructures for both production and dissemination purposes.

2. GATE (Geant4 application for Tomographic emission)

GATE (Geant4 application for Tomographic emission) is a C++ platform based on the Monte Carlo Geant4 software. It has been typically designed to model nuclear medicine

[1] **E**-Infrastructure Shared Between **E**urope and **L**atin **A**merica: http://www.eu-eela.org/

applications, such as PET and SPECT within the OpenGATE collaboration[5]. Its functionalities are combined to its ease of use make this platform also adequate for radiotherapy and brachytherapy treatment planning.

However, Monte Carlo simulations are computational-intensive, preventing hospitals and clinical centres from using them for daily practice. As a result, the objective of GATE is to use the Grid environment to reduce the computing time of Monte Carlo simulations in order to provide a higher accuracy in a reasonable period of time.

Nine Cuban centres are currently testing, as users the results of the simulation of radiotherapy treatments using realistic models that GATE provides. The interest of this community is around two main oncological problems:

- Thyroid Cancer is endemic in many areas of Cuba[1], being the diseases of thyroids one of the 5 main causes of Endocrinology treatments.
- Treatment of metastasis with P-32[2]. Brachytherapy using P32 isotopes is a procedure which is revealing very good results in Cuba. Improving the knowledge on the doses captured from the different tissues by accurate simulation is a key issue.

The main benefit of using the Grid is that it has enable medical users to access realistic Monte-Carlo simulation for their research in radiotherapy planning. Without the EELA Grid, medical users are not provided of enough computational resources to deal with the large requirements that this processing has. Grid will increase the performance to the application, but in this case it will even be more important, since it is an enabling technology opening the doors to a new range of applications.

All the centres from Cuba currently testing this application bring to the EELA community around 90 cases per month.

3. Wide in Silico Docking of Malaria (WISDOM)

The objective of WISDOM is the proposition of new inhibitors for a family of proteins produced by Plasmodium Falciparum. This protozoan parasite causes malaria and affects around 300 million people and more than 4 thousand people die daily in the world. Drug resistance has emerged for all classes of antimalarials except artemisinins. The main reason is that the available drugs focus on a limited number of biological targets, producing a cross-resistance to antimalarials. There is a consensus that substantial scientific effort is needed to identify new targets for antimalarials. The main problem is that the development of new drugs with new targets is a costly and lengthy process, and the economic profit is not clear for the drug manufacturers.

This application consists on the deployment of a high throughput virtual screening platform in the perspective of in silico drug discovery for neglected diseases. The WISDOM[1] platform performs a High-Throughput virtual Docking of million of chemical compounds available in the databases of ligands to several targets of Plasmepsin[6].

The interest of the EELA partners is basically centred in three actions:

- The study of new targets for new parasitory diseases. Such as Dengue as this endemic affects various countries in LA and fight against it has been very strong.

[1]This platform is being jointly developed by the SIMDAT, SwissBioGRI, Swiss Institute of Bioinformatics, the INSTRUIRE regional grid in Auvergne and the CampusGRID

Notwithstanding that several regions in LA communities are free of dengue, many other suffer from periodic epidemics. These targets will be added to the ones selected in the Data Challenge to maximise the exploitation of the resources.

- The selection of new targets for the malaria. This include new research lines on drug identification different from the ones selected in the WISDOM studies, which is more interesting for the LA communities, such as the DHFR protein and their chlorate derivatives.
- The contribution with resources for the WISDOM Data Challenge.

4. Basic Local Alignment Searching (BLAST)

One of the most important efforts on the analysis of the genome is the study of the functionality of the different genes and regions. Sequence alignments provide a powerful way to compare novel sequences with previously characterized genes. Both functional and evolutionary information can be inferred from well designed queries and alignments. The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

This process of finding homologous of sequences is a very computationally-intensive process. The size of the databases currently available (Non-redundant Gene Bank, SwissProt, etc.) increases daily, reaching the size of more than a gigabyte. Searching alignment of a single sequence is not a costly task, but normally, thousands of sequences are searched at the same time.

The biocomputing community usually relies on either local installations or public servers, such as the NCBI or the gPS@, but the limitations on the number of simultaneous queries make this environment inefficient for large tests. Moreover, since the databases are periodically updated, it will be convenient to periodically update the results of previous studies. Thus, the availability of an independent Grid-enabled version integrated on the Bioinformatics Portal of the Universidad de los Andes will provide registered users with results in a shorter time. A Grid service for the execution of BLAST on large sets of sequences has been developed at the UPV[3].

BLAST is being used for searching similar sequences and inferring their function in parasite disesases presented in Venezuela, such as the Leishmaniasis (mainly Mexican Leishmania), Chagas (mainly Trypanosoma Cruzi) and Malaria (mainly Plasmodium vivax). The use of Grids will enable increasing the number of fragments to be analysed and the periodical updating of this information. Moreover, the availability of larger scale computation will enable the researchers on performing evolutionary studies.

The increase on the capabilities on performance of the bioinformatics portal of the Universidad de los Andes through the use of the EELA Grid has brought more mature users who access this portal to use more powerful computational resources than those available on their centres.

5. Phylogeny (MrBayes)

A phylogeny is a reconstruction of the evolutionary history of a group of organisms. Phylogenies are used throughout the life sciences, as they offer a structure around which

to organize the knowledge and data accumulated by researchers. Computational phylogenetics has been a rich area for algorithm design over the last 15 years. The inference of phylogenies with computational methods is widely used in medical and biological research and has many important applications, such as gene function prediction, drug discovery and conservation biology [4]. The most commonly used methods to infer phylogenies include cladistics, phenetics, maximum likelihood, and MCMC-based Bayesian inference. These last two depend upon a mathematical model describing the evolution of characters observed in the species included, and are usually used for molecular phylogeny where the characters are aligned nucleotide or amino acid sequences.

The complexity of large-scale phylogeny studies, such as the "Tree of Life" project, aiming at all organisms on the Earth, represents a true computational grand challenge. Due to the nature of Bayesian inference, the simulation can be prone to entrapment in local optima. To overcome local optima and achieve better estimation, the MrBayes program has to run for millions of iterations (generations) which require a large amount of computation time. For multiple sessions with different models or parameters, it will take a very long time before the results can be analyzed and summarized. The phylogenetic tools are widely demanded by the Latin America bioinformatics community.

A Grid service for the parallelised version of MrBayes application is currently being developed and a simple interface will be deployed on the bioinformatics portal of Universidad de los Andes. This Grid-enabled service will make use of EELA resources to run phylogenetic studies at high performance.

6. Conclusion

The EELA e-infrastructure permits various collaborative groups in Latin America a more powerful computational resources than those available on their centers. This achieves that the lines of investigation stated in the document can be feasible as the computational requirements for them can be met. As well of creating a network where Latin American researches can participate in European initiatives and vice versa reducing the digital gap.

The EELA project currently has four biomedical pilot applications running, two from European initiatives (EGEE) and other coming from Latin American research groups. This project even though its in early phases will intend to bring more research lines into its e-infrastructure as well as creating a bigger network of collaborative partners.

References

[1] D. Navarro *"Epidemiología de las enfermedades del tiroides en Cuba"*, Rev Cubana Endocrinol 2004;15.

[2] J. Alert, J. Jiménez, *"Tendencias del tratamiento radiante en los tumores del sistema nervioso central"*, Rev Cubana Med 2004; 43(2-3).

[3] Gabriel Aparicio, Stefan Götz, Ana Conesa, J. Damian Segrelles Quilis, Ignacio Blanquer, J. Miguel García, Vicente Hernández, *"Blast2GO goes Grid: Developing a Grid-Enabled Prototype for Functional Genomics Analysis"*, Proceedings of the HealthGrid 2006 Conference.

[4] K. Lesheng , *"Phylogenetic Inference Using Parallel Version of MrBayes"*

[5] S. Jan, G. Santin, D. Strul, S. Staelens, *"GATE: a simulation toolkit for PET and SPECT"*, submitted to Phys. Med. Biol.

[6] V. Breton, *"Grid added value to fight neglected diseases"*, Wisdom Open Day