

Introducing Biotechvana Bioinformatics

Llorens, C.^{1,2}, Futami, R.², Muñoz-Pomer, A.^{2,3}, Vicente-Ripollés, M.^{2,3} and Moya, A.^{1,4}

1- Instituto Cavanilles de Biodiversitat i Biologia Evolutiva Universitat de València, Spain

2- Biotechvana, Valencia, Spain

3- Departament de Sistemes Informàtics i Computació Universitat Politècnica de València

4- CIBER de Epidemiología y Salud Pública (CIBERESP), Spain

Corresponding author: carlos.llorens@biotechvana.com

Availability: Available online February 1, 2008 at <http://biotechvana.uv.es>

In this paper, we introduce Biotechvana Bioinformatics; a self-sustaining initiative focusing on software advances, computational resources, and database utilities in the management of biological data. The platform facilitates access to an in-progress collection of tools presented through an online catalogue organized as an electronic journal, where tool manuals and other resources are distributed by sections presented as papers.

Keywords: Bioinformatics | Computational biology

INTRODUCTION

Since Hogeweg and Hesper used in 1978 the term “Bioinformatics” to refer to the process of information integration in biological systems (1), advances in molecular biology and genome sequencing have led to an impressive growth in the availability of biological information. With the also recent emergence and worldwide implantation of the Internet, the scientific community took advantage of Information Sciences in order to design new online databases and software tools to process, analyze and classify biological information. The term bioinformatics was then taken by the scientific community to collectively describe all advances involving algorithms and computational methodologies used to solve problems generated by the management and analysis of biological data. Note the difference with the term “Computational Biology” that is usually adopted to refer to the use of computers when investigating specific biological problems with the aim to enhance knowledge.

Progression of genomics and proteomics has also led to a significant increase of investigations focusing on the study of mobile genetic elements and the emergence of novel gene functions related to them. Most of these efforts have revealed that mobile genetic elements are more widely distributed in the genomes of eukaryotes than previously thought. They likely played an important role in the evolution of the complexity in life (2-8), and it is now accepted that the evolution from prokaryotes to eukaryotes was likely accompanied by changes in the nuclear genome, including expansions in size and number of introns and proliferation of mobile genetic elements (2).

We are particularly interested in the biological diversity and evolution of mobile genetic elements, and the evolution-

nary impact and diseases they cause in living organisms. With the aim to contribute in this area we have recently created the Gypsy Database (GyDB) of mobile genetic elements (9). This project is a long-term research focusing on the phylogenetic classification and relationships of mobile genetic elements and related nonviral proteins. The GyDB is an initiative among students and researchers of the University of Valencia and other institutions. The maintenance, growth and improvement of this infrastructure demand a significant effort in both human and technical resources. To indirectly guarantee our editorial independence we have created Biotechvana Bioinformatics (BB), which we introduce here; it is a bioinformatics platform available to other authors and industrial researchers interested in the development of original advances in Biology, Evolution, and Biomedicine.

OVERVIEW

The whole BB platform is implemented in a LAMP environment (Linux, Apache, MySQL and PHP). It is divided into four sections, arranged as follows:

Software. We use this section to introduce software tools. Currently the BB software collection offers a tool distributed under a closed source license, *Phylograph*, and the open-sourced *Checkalign*. *Phylograph* is a multi-function tree editor particularly indicated for large trees. *Checkalign* is a logomaker tool. You may also access this tool as an online public server at URL 1.

Database utilities. In this section we facilitate the *GyDB Package*, which comprises web-oriented solutions developed to build the GyDB along with other database utilities. For instance, the *Biotechvana Search Engine* is a cross-platform customizable engine to search web sites, or the *Biotechvana Queue Manager*: a server-overload preventing script.

Scripts. A collection of web-based utilities. The *Alignment Format Converter* script allows users to obtain various alignment formats in a single step from the input of one sequence alignment. *Join Alignments* is a script that concatenates several alignments into a single one. *RMXSC* is a PHP script that allows users to export files containing bibliographic data in *Reference Manager* XML format into a MySQL database.

Computational resources. In this section we facilitate access to the *GyDB Collection*: a repository of non-redundant multiple alignments, hidden Markov model (HMM) profiles (10) and majority-rule consensus (MRC) sequences based on all protein products encoded by *Ty3/Gypsy* and *Retroviridae* LTR retroelements and related nonviral proteins. The *GyDB Collection* is available at GyDB (9). In essence, this means

that the offered material can be used without restrictions if authors are properly cited. In this section we also make a computational cluster available for users that require computational resources to run analyses based on large data sets.

ADDITIONAL SERVICES

We maintain all resources via a permanent support service. We are also receptive to users' feedback as a peer-review mechanism which helps us improve the tools. The goal behind this initiative is similar to other internet worldwide actions where users can upload and download tools, scenarios, algorithms, exchange ideas, etc.

CITING

If you would like to refer to any resource provided by BB in your investigation, please do so by citing the latest version and section as it appears in its associated resource document in the BB Collection; an example follows:

Llorens, C., Futami, R., Vicente-Ripolles, M., and Moya, A. (2008) The CheckAlign logo-maker application in analyses of both gapped and ungapped DNA and protein alignments. In *Biotechvana Bioinformatics 2008*, Biotechvana, Valencia. Software: CheckAlign

CONCLUDING REMARKS

The BB collection is an online platform that furthers the progress of science through the continuous design of tools in computational biology and bioinformatics. We periodically re-edit the BB collection, online and in high-resolution ready-to-print formats. The collection is an initiative open to other authors to whom we encourage to upload their own tools. We are pleased to share algorithms, projects and ideas with other researchers interested in the area. Processed biological information, services, printed material and/or electronic documentation are distributed under the terms of the Creative Commons Attribution license (URL 2). This means that this type of material can be used without restrictions as long as authors are properly cited. For downloading the software and to have full access to all resources users are invited to agree to an annual subscription plan, to meet the expenses, which allows an unlimited number of licenses. If you are interested in more details, contact us at URL 3.

LITERATURE

1. Hogeweg, P. and Hesper, B. (1978) *Comput. Biol. Med.*, **8**, 319-327.
2. Lynch, M. and Conery, J.S. (2003) *Science*, **302**, 1401-1404.
3. Llorens, C. and Marin, I. (2001) *Mol. Biol. Evol.*, **18**, 1597-1600.
4. Ganko, E.W., Bhattacharjee, V., Schliekelman, P. and McDonald, J.F. (2003) *Mol. Biol. Evol.*, **20**, 1925-1931.
5. Brandt, J., Schrauth, S., Veith, A.M., Froschauer, A., Haneke, T., Schultheis, C., Gessler, M., Leimeister, C. and Volf, J.N. (2005) *Gene*, **345**, 101-111.
6. Jurka, J., Kapitonov, V.V., Kohany, O. and Jurka, M.V. (2007) *Annu. Rev. Genomics Hum. Genet.*, **8**, 241-259.
7. Volf, J.N. (2006) *Bioessays*, **28**, 913-922.
8. Kazazian, H.H., Jr. (2004) *Science*, **303**, 1626-1632.
9. Llorens, C., Futami, R., Bezemer, D. and Moya, A. (2008) *Nucleic Acids Research (NAR)*, **36** (Database-Issue):38-46.
10. Eddy, S.R. (1998) *Bioinformatics*, **14**, 755-763.

ACKNOWLEDGMENTS

We thank Rachel Epstein for language revision and Javi Ortiz and Isaac Fernández, Unitat de Bioinformàtica, Servei Central de Suport a la Investigació Experimental (SCSIE) at UVEG for technical support. Biotechvana Bioinformatics has been awarded the NOVA 2006 by IMPIVA and Conselleria d'Empresa, Universitat i Ciència of Valencia. The research has been partly supported by grants IMCBTA/2005/45, IMIDTD/2006/158 and IMIDTD/2007/33 from IMPIVA, and by grant BFU2005-00503 from MEC to AM.

URLS

1. **CheckAlign server:** <http://gydb.uv.es/servers/checkAlign>
2. **Creative Commons Attribution License:** <http://creativecommons.org/licenses/by/2.0>
3. **Contact Web Site:** <http://biotechvana.com/loader.php?section=contents&page=contact>

SPONSORS



VNIVERSITAT  DE VALÈNCIA



UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
Una manera de hacer Europa