Bittechvana

The GyDB collection is now publicly available within the Gypsy Database of Mobile Genetic Elements

Laura Covelli¹, Alfonso Muñoz-Pomer^{1,2}, Laura Domínguez-Escribà¹, Ricardo Futami¹, Guillermo Pablo

Bernet¹, Andrés Moya^{3,4}, Carlos Llorens¹

1. Biotechvana, Parc Cientific de la Universitat de València

2. Departamento de Sistemas Informáticos y Computación (DSIC), Universitat Po-litècnica de València

3. Unidad Mixta de Investigación en Genómica y Salud del Centro Superior de In-vestigación en Salud Pública (CSISP)-Universitat de València (Instituto Cavanilles de Biodiversidad y Biología Evolutiva)

4. CIBER en Epidemiologia y Salud Pública (CIBEResp)

Corresponding author: carlos.llorens@biotechvana.com

Availability: Available online February 28th, 2011 at [URL 1]

The GyDB (Gypsy Database) collection is a nonredundant repository of multiple alignments, hidden Markov model profiles, and majorityrule consensus sequences. The collection is based on all currently known protein domains of the distinct mobile genetic elements and related host genes classified at "Gypsy Database of mobile genetic elements". Alignments are available in six formats: FASTA, PIR, MSF, Stockholm, Clustal and Phylip. Hidden Markov model profiles and consensus sequences are constructed based on each protein domain consensus accepted per monophyletic group of classified MGEs and protein domains. The GyDB collection was originally launched as a resource of Biotechvana Bioinformatics that is related to, but it is independent from, the Gypsy Database. This relationship has recently motivated us to finally deposit the GyDB collection within the Gypsy database, where the resource is now publicly accessible as a permanent section.

Keywords: Mobile Genetic Elements | Multiple alignments | HMM profiles| MRC sequences

OVERVIEW

The Gypsy Database (GyDB) is an ongoing project [1,2] devoted to the evolutionary dynamics of viruses and mobile genetic elements (MGEs), which was launched in 2008. The GyDB is a highly informative database established within an evolutionary context of classification, where one piece of research delivers one conclusion that drives individuals towards another goal. The most captivating aspect of this project is that a share of our efforts are dedicated to the interpretation of analyses, paying particular attention to non-redundant MGEs displaying a certain degree of phylogenetic distance, and investigating how they can be collectively related in terms of protein domain architecture with other lineages and elements. In particular, we evaluate the phylogenetic signal of the distinct classified elements by creating alignments,

Hidden Markov Model (HMM) profiles [3] and majority rule consensus (MRC) sequences, per lineage and protein domain. The whole material is implemented in a database repository we call "GyDB Collection" [4,5]. This repository was originally launched as a resource of Biotechvana Bioinformatics [6]. However its direct relationship with the GyDB project has motivated us to move the GyDB collection from Biotechvana Bioinformatics to the GyDB. In this paper we formally communicate that the GyDB collection is now a publicly available section, as described in the release 2.0 of the GyDB project [1]. As shown in Figure 1, the collection is organized in three categories: multiple alignments, HMM profiles and MRC sequences. Multiple alignments are available in six formats: FASTA, PIR, MSF, Stockholm, Clustal and Phylip. Users have the possibility to consult the different sections of the GyDB collection or download it.

ACKNOWLEDGMENTS

GyDB Collection has been partly supported by grants 17092008 from ENISA (Empresa Nacional de Innovacion S.A), IDI-20100007 from CDTI (Centro de Desarrollo Tecnológico Industrial) and PTQ-09-01-00020 and PTQ-09-01-00670 from MICINN (Ministerio de Ciencia e Innovación).

Funding to pay the Open Access publication charges for this article was provided by the University of Valencia.

LITERATURE

1. Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D et al.: The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res* 2011, **39**: D70-D74.

2. Llorens C, Futami R, Bezemer D, Moya A: The Gypsy Database (GyDB) of Mobile Genetic Elements. *Nucleic Acids Res* 2008, **36**: 38-46.

3. Eddy SR: Profile hidden Markov models. *Bioinformatics* 1998, 14: 755-763.

4. Llorens C, Muñoz-Pomer A, Futami R, Moya A: The GyDB Collection of Viral and Mobile Genetic Element Models. In *Biotechvana Bioinformatics*. Biotechvana, Valencia; 2009:CR: GyDB Collection.

5. Llorens C, Futami R, Moya A: The GyDB collection: Ty3/Gypsy and Retroviridae LTR retroelements and related nonviral proteins. In *Biotechvana Bioinformatics*. Biotechvana, Valencia; 2008:CR: GyDB Collection.

6. Llorens C., Futami R., Muñoz-Pomer A., Vicente-Ripolles M, Moya A.: Introducing Biotechvana Bioinformatics. *Biotechvana Bioinformatics* 2008, intro.

URL

1. GyDB Collection: http://gydb.org/index.php/Collection_alignments



Figure 1. Screenshot of the GyDB Collection at GyDB