

An integrated ontological knowledge base about orthologous genes and proteins

J. A. Miñarro-Gimenez¹, Marisa Madrid², J. T. Fernandez-Breis¹

¹Departamento de Informatica y Sistemas, Universidad de Murcia, Spain, emails: {jose.minyarro, jfernand}@um.es

²Cell Division Group, Paterson Institute for Cancer Research, University of Manchester, UK, email:Mmadrid@picr.man.ac.uk

Abstract

The integration of biomedical resources is necessary due to the amount of heterogeneous information continuously being generated. In this work, we address the integration of existing information about orthologous genes and proteins. The process followed in this work consists of three main steps: (1) construction of the ontology representing the biological knowledge of the resources; (2) definition of the mappings between the ontologies and the resources, which are needed to support the data integration process; and (3) integration, through which an ontological knowledge base is created. The ontology was implemented in OWL, containing 52 concepts, 9 object properties and 2 data type properties, cardinality and disjointness constraints. As a result, a semantic resource, which integrates the resources used by YOGY[1], has been obtained, containing approximately 1168000 orthologous genes, 956000 proteins connected to genes, and 114000 orthologous clusters. The development of this integrated ontological repository provides a series of advantages: (1) more powerful and efficient usage of the information; (2) the integration process allows for removing redundancy, so the information improves on precision and quality as well as reduces time and effort for researchers; (3) the repository can be integrated with different tools using the same ontology, and the knowledge can easily be enriched and extended by reusing other bio-ontologies.

ACKNOWLEDGEMENTS

Jose Antonio Miñarro is supported by the Fundacion Seneca. Marisa Madrid is supported by an EMBO Long-Term Fellowship. Thanks to the Spanish Ministry for Science and Education through the project TSI2007-66575-C02-02, and to Valerie Wood and Jurg Bähler for their support and their valuable comments.

References

1. Penkett, C., Morris, J., Wood, V., Bähler, J.: Yogy: a web-based, integrated database to retrieve protein orthologs and associated gene ontology terms. *Nucleic Acids Res* **34** (2006) W330–4