

Plant-Pathogen Interactions Ontology (PPIO)

Alejandro Rodríguez Iglesias ^{a,*}, Mikel Egaña Aranguren ^a Alejandro Rodríguez González ^a
Mark D. Wilkinson ^a

^a *Biological Informatics Group, Centre for Plant Biotechnology and Genomics (CBGP), Technical University of Madrid (UPM), Spain*

Abstract. Plant-pathogen interactions are an important knowledge domain within plant biology and biotechnology, both scientifically and in economic terms. Unlike other knowledge domains within life sciences, however, semantic technologies have not been used extensively to codify it; therefore, there is a lack of axiomatic models amenable to automated integration and inference. We present the Plant-Pathogen Interactions Ontology (PPIO), a first step towards the axiomatization of plant-pathogen interactions knowledge. PPIO encourages consistent annotation and supports both query and inference.

Keywords: Plant pathogenic bacteria, Ontologies, Semantic Web, PPIO

1. Introduction

Many bacterial genera are responsible for causing diseases in a wide range of plant species, and if the different levels of the host defense barriers are overcome by the pathogen, the infection process can ultimately lead to the death of the plant. This phenomenon can affect crop yield, which automatically will translate into considerable economic loss. On the other hand, in the context of challenges such as feeding a growing world population, and the trend of emerging economies to consume more meat, where animal feed increases the consumption per capita of available food-crops, it is crucial that we identify ways to increase the productivity of crop fields. Therefore, being able to accurately record, explore, and query biological data regarding plant-pathogen interactions is a necessary step towards the improvement and protection of these important plant species. This issue continues being subject of intensive research worldwide. This is documented in hundreds of articles that focus on the biological consequences and the mechanisms of pathogenic bacteria interactions with their hosts.

Certain semantics-oriented projects are particularly noteworthy as a result of their success in using semantics to aid in automated data integration from

non-collaborating resources, like the OBO foundry, Bio2RDF and the W3C Semantic Web for Health Care and Life Sciences Interest Group¹. In the context of plant biotechnology and phytopathology, however, semantic technologies have been applied to only a limited number of domain resources. For example, the Plant Ontology Consortium developed the Plant Ontology² platform to describe the anatomy, morphology and developmental stages of plants. The Plant Trait Ontology, related to the Gramene project³, defines a vocabulary for describing the specific appearance or qualities of various plant anatomical structures. The Plant Disease Ontology (IDOPlant), was modeled after the human Infectious Diseases Ontology (IDO) to serve as a reference ontology that contained and covered any plant infectious disease. Apart from IDOPlant, there is another contribution directly related to the plant-pathogenic bacteria area, and is the GO extension for description of the Type III Effectors, which is part of the Plant-Associated Microbe Gene Ontology project (PAMGO).

As just discussed, Semantic technologies have proven extremely effective and have widely adopted in other branches of the life sciences community. However,

*Corresponding author. Email:alejandroriglesias@gmail.com

¹<http://www.w3.org/blog/hcls/>

²<http://www.plantontology.org/>

³<http://www.gramene.org>

they have not been extensively applied to the knowledge domain of plant-pathogenic bacteria to date. Despite the clear need for automation of data annotation, categorization, integration, interpretation, and quality-control, the data and knowledge in this domain remains largely computationally opaque, limiting the utility and usability of these important datasets. Here we present the Plant Pathogen Interactions Ontology (PPIO), an ontological platform that integrates data related to plant pathology and physiology in the context of pathogenic interactions. The aim of PPIO is to leverage this combined knowledge to assist in the interpretation of the phenotypic responses that result from plant-pathogenic bacteria interactions. This platform offers an axiomatic scaffold into which biologically relevant data can be embedded in a precise and computationally-effective accessible manner. PPIO will make it possible for users to pose, and answer, questions like the following ones and obtain a meaningful answer:

1. Is tomato plant susceptible to the attack of *Pseudomonas syringae* pv. *tomato* DC3000?
2. Does a high humidity and and low temperature rates favour the development of *Pectobacterium carotovorum* subsp. *carotovorum*?
3. What is the phenotype of the disease produced by *Dickeya dadantii*?

2. Modelling

2.1. Desing principles

3. Creation methodology

Numerical IDs + label.

The development of PPIO is automated as much as possible. Once the basic structure is set, some parts of the ontology are produced programmatically by introducing the ontology in a tailored Galaxy [2] workflow (picture of workflow).

OWL Punning⁴ is used in the X hierarchy of PPIO to represent X classes both as OWL classes and individuals, in order to link pathogens, which are individuals, with symptoms?. This is achieved by defining an Ontol-

ogy Pre Procesoor Language (OPPL)⁵ script (picture) and executing it via OPPL-Galaxy [1].

⁴<http://www.w3.org/TR/owl2-new-features/>
Punning

⁵<http://oppl.sf.net>

The organism taxa hierarchy is produced by the Galaxy tool NCBITaxonomy2OWL: NCBITaxonomy2OWL gets the user-defined taxa from the NCBO web service (URL) and injects them in the ontology, respecting the original hierarchy and adding each taxa with an ontology (REF) URI (eg) (ref github repo)

By using Galaxy, the specifi workflow we need is defined once and we can execute it at will, each time a new release is set, or also if new tools are needed. The galaxy workflow can be reproduced at biordf.org:8090 with any ontology and OPPL script.

TODO: we need to be able to resolve not only the whole ontology (oclc.purl.org/PPIO), but specific entities (oclc.purl.org/PPIO000023). How? stardog linked data? but then, the redirection is always done with oclc.purl.org!

4. Discussion (comparison with other ontologies on the same topic, pointers to existing applications or use-case experiments)

Acknowledgements

Alejandro Rodríguez Iglesias and Alejandro Rodríguez González are funded by the Isaac Peral Programme. Mark D. Wilkinson and Mikel Egaña Aranguren are funded by the Marie Curie-COFUND Programme (FP7) of the EU.

References

- [1] M. E. Aranguren, J. T. F. Breis, E. Antezana, C. Mungall, A. R. González, and M. Wilkinson. OPPL-Galaxy, a Galaxy tool for enhancing ontology exploitation as part of bioinformatics workflows. *Journal of Biomedical Semantics*, 4(1):2+, 2013.
- [2] J. Goecks, A. Nekrutenko, J. Taylor, and Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*, 11(8):R86+, 2010.