The Ontology of Microbial Phenotypes (OMP): A Precomposed Ontology Based on Cross Products from Multiple External Ontologies that is Used for Guiding Microbial Phenotype Annotation

Marcus Chibucos[1], Adrienne Zweifel[2], Deborah Siegele[2], Peter Uetz[3,4], Michelle Giglio[1], James Hu[2]

[1]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA
[2]Texas A&M University, College Station, TX, USA
[3]University of Delaware, Newark, DE, USA
[4]Proteros Biostructures, Gaithersburg, MD, USA and Martinsried, Germany

Abstract. The Ontology of Microbial Phenotypes (OMP) is being developed to standardize capture of phenotypic information, including both processes and physical characteristics, from microbes. The OMP team comprises ontologists, microbiologists, and annotators, and ontology development is being performed in conjunction with the development of a wiki designed for annotation capture. Term development is being guided by following, to as great an extent as possible, the structure of existing ontologies. All OMP terms have Aristotelian definitions, and, when appropriate, they have genus-differentia cross products composed of terms from external ontologies. Initially, OMP is being used to annotate the prokaryotic model organism *Escherichia coli*. Eventually we anticipate that diverse user groups will employ OMP for standardized annotation of various microbial phenotypes, much in the same way that the Gene Ontology has standardized the annotation of gene products. Definitions of phenotypes and links to the original literature will facilitate the experimental characterization of phenotypes.

Keywords: annotation capture, bacterial phenotype, Aristotelian definition, cross product, *Escherichia coli*, GONUTS, microbe, microbial phenotype, phenotype annotation, wiki.

1 Introduction

Microbial Phenotypes and the Need for an Ontology

A phenotype is the expression of a genotype (i.e. the full genetic complement of an organism) in a given environment. For example, eye color, number of seeds per pod, and coat color are phenotypic traits that can be observed in flies, lupines, and ponies, respectively. Within an individual organism, both changes in genetic makeup, such as from bacterial conjugation, and variation in gene expression can result in different phenotypes under similar environmental conditions. Conversely, environmental variation can lead to different outcomes for genetically identical organisms, through variable gene expression. Myriad genetically and taxonomically diverse microbes exhibit countless variability in their morphological and physiological traits, both within and among species. Oftentimes these result in unique and exquisite manifestations, such as the symbiosis between the bioluminescent *Vibrio fischeri* bacterium and its squid host *Euprymna scolopes*. Characterization of phenotypes is critically important for medical microbial identification, and many unique biotechnological applications of microbes are rooted in phenotypes. Genetic manipulation with associated phenotypic characterization remains an important tool for determining protein function in microorganisms amenable to manipulation, such as *Escherichia coli*. To facilitate research in all of these areas, we are developing the Ontology of Microbial Phenotypes to allow for standardized capture of essential phenotypic information.

2 Previous Work

Manual versus Cross Product Terms

Previously, we explored two parallel approaches for building the Ontology of Microbial Phenotypes (OMP) [1] using the ontology editor OBO-Edit [2] and a custom script. We read 100 papers involving metabolic phenotypes and from these we identified 40 microbial phenotypes. We manually created an ontology comprising five super classes to represent those phenotypes. Separately, we created an ontology of cross products between selected PATO [3] terms and two nodes from the Gene Ontology [4], *cellular carbohydrate metabolic process* (GO:0044262) and *cellular amino acid metabolic process* (GO:0006520), which encompassed those phenotypes. Both approaches had advantages. Creating automated cross products was faster; manual term generation often reflected better how a biologist or annotator might think of a phenotype. Both highlighted the importance of creating synonyms and manually wording English definitions in a colloquial syntax, while adhering to strict equivalence guidelines.

Structure of the Ontology of Microbial Phenotypes

The first version of the Ontology of Microbial Phenotypes comprising 252 terms was released in June, 2011, and can be downloaded from SourceForge [5]. As its root class, OMP has *microbial phenotype* (OMP:0000000), defined as "the manifestation of a microbe's genotype in an environment." Descended from the root are terms that describe various attributes of microbial phenotypes including: cell arrangement, cell staining, cellular development, cellular morphology, metabolism, motility, multiorganism interactions, and response to stimulus.

Physical Objects and Processes. Phenotypes described by OMP include both physical objects and processes; descendants of the root *microbial phenotype* address both morphological and physiological traits. For example, *absence of flagellum* (OMP:0000030) is a descendant of *cellular morphology phenotype* (OMP:0000071), which is defined as

"a microbial phenotype, where the trait in question is the form and structure of the cell." However, all of the terms that describe the process of motility are descended from a distinct class *motility phenotype* (OMP:0000001), defined as "a microbial phenotype where the trait in question is the self-propelled movement of a cell from one location to another."

Relative versus Absolute Phenotypes. Some phenotypes described by OMP are relative. A process might be altered relative to how a typical organism performs that process, or an organism might not possess a cell part that it would typically possess. For example, a type of microbe that usually possesses a flagellum, but which does not possess a flagellum due to a mutation, is described by the relative term *loss of flagellum* (OMP:0000032). In contrast, some OMP terms describe non-relative characteristics that are inherent in an organism. For example, a type of microbe that does not naturally produce a flagellum would be described by *non-flagellated* (OMP:0000019).

Aristotelian Definitions. All terms have Aristotelian (genus-differentia) definitions of the form "B is an A that C's." All terms composed with cross products have cross product definitions, which describe a quality that inheres in a thing, whether a process or an object [6]. For example, the OMP term *abolished motility* (OMP:0000044) has the cross product definition "abolished inheres in cell motility" constructed from the PATO term *abolished* (PATO:0001508) and the GO term *cell motility* (GO:0048870).

3 Wiki for Annotation Capture

We are implementing a wiki modeled on Gene Ontology Normal Usage Tracking System (GONUTS) [7], our wiki for Gene Ontology, for exploring the ontology, adding usage notes to terms, and making phenotype annotations. We use a custom table-editing extension to provide structured data entry and data-mining capabilities. We envision annotation of different types of entities: taxonomic entities such as species and strains, mutant phenotypes, and phenotype predictions.

4 Downloads and Becoming Involved

Two versions of the ontology are available for download; one contains cross products from other ontologies, and the other is a streamlined version with no cross products. Both versions have database cross references, where appropriate. The OMP team welcomes community involvement in the development and application of OMP. We maintain a wiki [8] to facilitate discussion of ontology and annotation related issues, and we have a Source Forge tracker [9] for term requests.

Acknowledgments

References

1. Giglio, M., Mungall, C., Uetz, P., Yin, L., Goll, J., Siegele, D., Chibucos, M., Hu, J.: Development of an Ontology of Microbial Phenotypes (OMP). Nature Precedings. (2009)

2. OBO-Edit, http://oboedit.org

3. Phenotypic Quality Ontology, http://obofoundry.org/wiki/index.php/PATO:Main_Page

4. The Gene Ontology, http://www.geneontology.org

5. SCM Repositories – microphenotypes, http://microphenotypes.svn.sourceforge.net

6. Mungall, C.J., Gkoutos, G.V., Smith, C.L., Haendel, M.A., Lewis, S.E., Ashburner, M.: Integrating Phenotype Ontologies across Multiple Species. Genome Biology. 11:R2 (2010)

7. Gene Ontology Normal Usage Tracking System (GONUTS), http://gowiki.tamu.edu/wiki/index.php/Main_Page

8. Ontology of Microbial Phenotypes wiki, http://microbialphenotypes.org/wiki/.

9. SourceForge site for the Ontology of Microbial Phenotypes, http://sourceforge.net/projects/microphenotypes