

# OGG-Bac: Ontological Representation of Orthologous Bacterial Genes

Meng Liu<sup>a,b</sup>, Hong-Yu Ou<sup>a</sup>, Yongqun He<sup>b</sup>

<sup>a</sup> State Key Laboratory of Microbial Metabolism, School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai, China. <sup>b</sup> University of Michigan Medical School, Ann Arbor, MI 48109, USA

## Abstract

Different from eukaryotic organisms that turn to have one set of gene annotations, bacteria are usually annotated with strain-specific genome annotations within which a large number of genes (e.g., DNA replication initiation factor *dnaA* gene) may be orthologs. To systematically integrate and analyze bacterial genes, we developed a bacterial branch of the Ontology of Genes and Genomes (OGG-Bac) with a focus on standardized orthologous gene representation and annotations.

## Keywords:

OGG; ortholog; bacteria gene

## Introduction

Eukaryotic organisms usually have one set of gene annotations for each species (e.g., human). For bacteria, each bacterial species turns to have decades and even hundreds of strains with sequenced genomes, and each genome is individually annotated. Therefore, it is very difficult to integrate the huge number of bacterial genes in the same species or at higher taxonomy levels.

Orthologous genes have evolved from a common ancestral gene. Using the OGG (an OBO Foundry library ontology) as a platform, an ontological representation of the hierarchy and properties of the bacterial orthologs at different taxonomic levels can facilitate the standardization and better understanding of the complex relations among various bacterial genes.

## Methods

OGG-Bac was aligned with the OGG-core. The web servers Ontobee (<http://www.ontobee.org/>) and Ontofox (<http://ontofox.hegroup.org>) were applied to check and import the related existing terms from reliable ontologies. The Protégé-OWL editor was used for manually ontology editing.

## Results

Figure 1 uses gene *dnaA* to illustrate how OGG-Bac represents different levels of bacterial orthologs. The *dnaA* gene encodes the DNA replication initiation factor in bacteria. It is the first gene in the synthetic minimal bacterial genomes (1) and conserved in various bacteria (e.g., *Brucella* and *E. coli*).

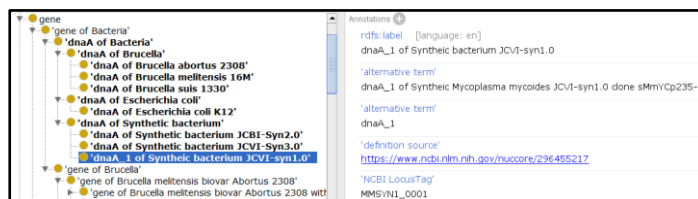


Figure 1. Representation of *dnaA* gene in OGG-Bacteria

Our representation strategy principles: (i) Use unique gene IDs for different levels of *dnaA*. (ii) Strain-specific genes are asserted based on their taxonomy hierarchy. (iii) For annotation, use the most common gene name as the gene label, and includes other gene labels (if different) as the alternative labels (Figure 1).

We are first representing the genes in the synthetic minimal bacterial genomes (1) and their orthologous genes in other bacteria.

## Conclusions

OGG-Bac was developed to ontologically represent and annotate bacterial genes based on the ortholog gene hierarchy at different taxonomical levels.

## Acknowledgments

National Key R&D Program of China [2017YFC1600100 to H.Y.O.]; ML was supported by a jointly funded Ph.D.-studentship of the China Scholarship Council and University of Michigan Medical School (Grant No. 201806230209).

## Address for correspondence

YH and HYO are the co-corresponding authors. Email addresses: [yongqunh@med.umich.edu](mailto:yongqunh@med.umich.edu) and [hyou@sjtu.edu.cn](mailto:hyou@sjtu.edu.cn).

## References

1. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., et al. (2016) Design and synthesis of a minimal bacterial genome. *Science*, 351, aad6253.