

FINAL TECHNICAL REPORT: US Department Of Energy Grant DE-SC0005011

November 3, 2017

Title: Gene Ontology Terms and Automated Annotation for Energy-Related Microbial Genomes

Website: <http://www.mengo.biochem.vt.edu/>

Program Officer: Dr. Pablo Rabinowicz, Office of Biological and Environmental Research
U.S. Department of Energy Office of Science

PIs: Biswarup Mukhopadhyay (final lead), Brett M. Tyler (early lead), Joao C. Setubal (interim lead) and T. M. Murali

Project Senior Personnel: Endang Purwantini and Trudy Torto-Alalibo

Institution: Virginia Polytechnic Institute and State University

Lead PI contact: e-mail: biswarup@vt.edu; Phone: 540 231 1219

Virginia Tech Sponsored Research Office Contact:

Nevada Davis, e-mail: ngdean@vt.edu; Phone: (540) 231 1427

Project period: May 17, 2010 - Aug 14, 2016

Non-funded collaborating institutions: DOE Joint Genome Institute (JGI), Walnut Creek, CA; Great Lakes Bioenergy Research Center (GLBRC), Madison, WI; BioEnergy Science Center (BESC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN; Joint BioEnergy Institute (JBEI), Berkeley, CA.

EXECUTIVE SUMMARY

Gene Ontology (GO) is one of the more widely used functional ontologies for annotating genes. The project developed 660 GO terms for describing energy-related microbial processes and filled the known gaps in this area of the GO system, and performed 179 gene annotations to showcase the utilities of these terms. It hosted a series of workshops and made presentations at key meetings to inform and train microbiologists on these terms and to receive inputs from the community members for the GO term generation efforts. The project developed a website for storing and displaying these terms and annotations (<http://www.mengo.biochem.vt.edu/>). The website also presents a set of relevant bioenergy production pathways displaying the associated GO terms. The outcome of the project was further disseminated through peer-reviewed publications and poster and seminar presentations.

PROJECT OBJECTIVES

1. To develop a set of GO terms [Microbial ENergy-related Gene Ontology (MENGO) terms] for annotating energy-related microbial genomes. (Fig. 1)

2. To develop a database and web interface for storing and displaying manual annotations.

The project also initiated an effort towards developing GO terms and Annotations for Synthetic Biological Processes (SYNGO).

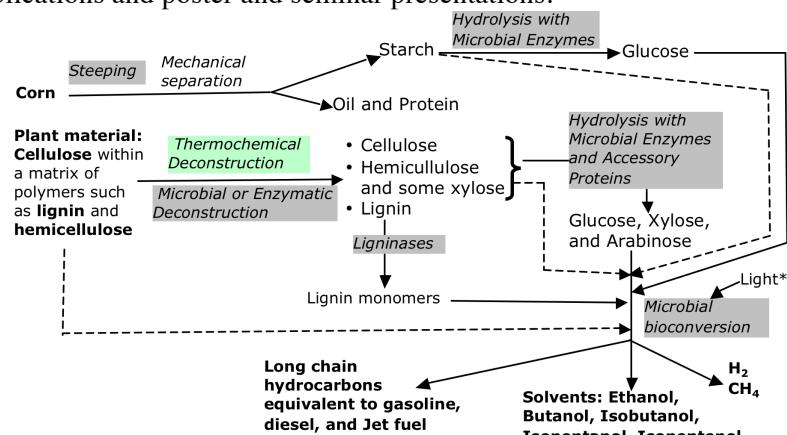


Fig. 1. Biofuel production from grains and lignocellulosic materials, existing and conceptual processes. The feed-stocks and biofuels are shown in bold. The dotted lines describe the consolidated bioprocesses (CBPs). The bioconversion processes and reagents are shown shaded in grey; the green shaded step is to be replaced with a biological process. The microbial processes could involve single or multiple organisms, naturally occurring or engineered variants. *Sunlight can be used as sole or supplementary energy source for biofuel production.

For outreach to the community and to derive benefit for the term development and annotation tasks from the community's expertise, the project conducted workshops at the Annual DOE Genomic Science Awardee Meetings, General Meetings of the American Society for Microbiology, the Annual User Meetings of the DOE Joint Genome Institute (DOE-JGI) and GLBRC, and had meetings with the collaborators. The project built a web interface under objective 2 to serve as MENGOfocal and to provide information about term development and upcoming and past meetings, and training materials.

ACCOMPLISHMENTS

Objective 1

Term generation:

The project generated 660 MENGOfocal terms for microbial metabolisms related to biofuel production and these have been incorporated into the Gene Ontology (GO) consortium database (<http://amigo.geneontology.org/>). These terms span all three ontologies in the GO: Molecular Function (87 terms), Biological Process (557 terms) and Cellular Component (16 terms).

Areas addressed in the term generation effort:

The term generation effort mainly concerned the following processes: Biomass Degradation (345 terms), Methanogenesis (110 terms), Ethanol Biosynthesis (11 terms), Butanol Biosynthesis (13 terms), Alkane Biosynthesis (31 terms), Alkene Biosynthesis (43 terms), Alkadiene Biosynthesis (14 terms), and Hydrogen Biosynthesis (14 terms).

Several existing terms were modified to provide definitions that are more relevant to bioenergy production and to make existing bioenergy-relevant terms visible to the community, we have added synonyms/alternative names that are widely used in the literature.

Ready presentation of MENGOfocal terms on metabolic pathways:

The project presented the MENGOfocal terms on associated microbial bioenergy production pathways for ready visualization. This presentation is key feature of the project website (<http://www.mengo.biochem.vt.edu/>) and publications (listed below).

Objective 2

Gene annotation:

The goal of this activity was to show the utility of the MENGOfocal terms. The project developed 179 gene annotations. To provide the high quality examples, only the genes with clear experimental evidences for their functions, such as those Inferred from Direct Assays [IDA] or Mutant Phenotypes [IMP]), were annotated and the annotations were performed manually.

Collaboration with the community and the outcome:

The project organized five workshops at the following venues: GLBRC (February 2011); Annual DOE Genomic Science Awardee Meetings (April 2011 and February 2012); Annual JGI User Meeting (March 2011 and March 2012). The project leader also visited the Computational Biology and Bioinformatics group at the ORNL for a discussion about the MENGOfocal project. The posters and seminars presented by the project, as listed below, also helped to gather inputs from the community members. All these interactions helped the project in establishing contacts with a broad range of bioenergy researchers and getting fresh ideas about term generation.

A website focused on the needs of the bioenergy research community:

The project developed a website (<http://www.mengo.biochem.vt.edu/>) for the following purposes.

(i) To display the project results:

- MENG0 terms
- Experimental evidence-based manual gene annotations
- Microbial bioenergy production pathways displaying the associated GO terms

(ii) To use it as an avenue for communication with the relevant research communities for the purpose of (a) bringing the project resources to their notice and (b) leveraging their expertise in the development of the MENG0 terms and gene annotation tasks, as well as for educating the community members in the use of GO terms in gene annotation.

The website has menus for the following:

- A list of MENG0 terms that have already been approved by the GO consortium and incorporated into the AMIGO database and additional terms that are either yet to be approved or underdevelopment or even do not fit the current scope of the GO consortium.

In addition to proving a comprehensive list, the site present process specific and ontology specific lists.

- A form that the community experts at various levels (undergraduate/graduate students, post-docs, and experienced researchers) could use to suggest new GO terms or bring processes/systems that need to be defined with new GO terms to our attention.
- A form under the “Gene Annotation” section that the community members could use to generate gene annotations for considerations by project.
- Various microbial metabolic pathways that are used for bioenergy production, with individual steps annotated with relevant GO terms and a link to AMIGO. This system allows ready visualization of the MENG0 terms and other relevant GO terms in the context of the associated biological processes and systems, which is very useful to the segments of the community that are not familiar with the concepts of Gene Ontology (GO) or microbiology/biochemistry.

The following pathways are under display at the site:

- Methane Biosynthesis (Methanogenesis) and related metabolisms, such as biosynthesis of Coenzyme F₄₂₀, Coenzyme M, Tetrahydromethanopterin, Methanofuran, Coenzyme B, and Coenzyme F₄₃₀
- Alcohol Biosynthesis - Ethanol, Isopropanol, Propanol (via dehydratase and keto-acid decarboxylase activities and via citramalate), Butanol, and Isobutanol
- Long-chain Alkane Biosynthesis
- Long-chain Alkene Biosynthesis
- Hydrogen Biosynthesis via Hydrogenase, Nitrogenase, and Formate Hydrogen Lyase (FHL) activities
- Fatty Acid Ester Biosynthesis (Fatty acid methyl ester; Fatty acid ethyl ester).

An Additional Product (SYNGO):

A review of the activities in the Bioenergy field and interactions with the community members as presented above, brought up a need for developing GO terms and Annotations for Synthetic

Biological Processes. Since such an endeavor does not fall within the scope of GO, which concerns only natural processes, the project byproducts along this line have been presented at the project website under a new title, SYNGO. The website documents 14 SYNGO terms, 6 Synthetic Gene Annotations and following two biosynthetic pathways:

- Butanol/Isobutanol Biosynthetic Pathway (from Glycine)
- Acetone/Butanol/Ethanol Biosynthetic Pathway

Peer-reviewed publications from the project

1. Purwantini, E., T. Torto-Alalibo, J. Lomax, J.C., Setubal, Tyler, B.M. and B. Mukhopadhyay. 2014. Genetic resources for methane production from biomass described with the Gene Ontology. *Front. Microbiol.* 5:634.
<https://www.frontiersin.org/articles/10.3389/fmicb.2014.00634/full>
2. Torto-Alalibo, T., E. Purwantini, E., J. Lomax, J.C., Setubal, B. Mukhopadhyay and B.M. Tyler. 2014. Genetic Resources for Advanced Biofuel Production Described with the Gene Ontology. *Front. Microbiol.* 5:528.
<https://www.frontiersin.org/articles/10.3389/fmicb.2014.00528/full>

Oral Presentations from the project

Biswarup Mukhopadhyay (presenter). June 2012. Hydrothermal vent biology of *Methanocaldococcus jannaschi* and development of sulfate reduction pathway, Bioinformatics Research Group, SRI International, Menlo Park, CA. (*Invited Seminar*)

Poster presentations from the project

1. E. Purwantini, T. Torto-Alalibo, J. C. Setubal, B. M. Tyler, B. Mukhopadhyay. Microbial Energy Processes Gene Ontology (MENGO): Development of terms and annotation for energy-related microbial genomes. 112th General Meeting of the American Society for Microbiology, June 16 - 19, 2012, San Francisco, CA.
2. E. Purwantini, T. Torto-Alalibo, J. C. Setubal, B. M. Tyler, and B. Mukhopadhyay. Gene Ontology terms describe biological production of methane. Seventh Annual Department of Energy Joint Genome Institute Users Meeting, March 20-22, 2012, Walnut Creek, CA.
3. T. Torto-Alalibo, E. Purwantini, J. C. Setubal, B. M. Tyler, and B. Mukhopadhyay. Microbial ENergy processes Gene Ontology (MENGO): New Gene Ontology terms describing microbial processes relevant for bioenergy. Seventh Annual Department of Energy Joint Genome Institute Users Meeting, March 20-22, 2012, Walnut Creek, CA.
4. T. Torto-Alalibo, E. Purwantini, J. C. Setubal, B. M. Tyler, and B. Mukhopadhyay. Microbial ENergy processes Gene Ontology (MENGO): New Gene Ontology terms describing microbial processes relevant for bioenergy, Department of Energy Genomic Science Awardees Meeting X. February 26-29, 2012, Bethesda, MD.
5. E. Purwantini, T. Torto-Alalibo, J. C. Setubal, B. M. Tyler, and B. Mukhopadhyay. Gene Ontology terms describe biological production of methane. Department of Energy Genomic Science Awardees Meeting X, February 26-29, 2012, Bethesda, MD.
6. T. Torto-Alalibo, E. Purwantini, B. Mukhopadhyay, B. M. Tyler and J. C. Setubal. Gene Ontology for microbial processes useful for bioenergy production. Department of Energy Genomic Science Awardees Meeting IX, April 10-13, 2011, Crystal City, VA.
7. T. Torto-Alalibo, E. Purwantini, B. Mukhopadhyay, B. M. Tyler and J. C. Setubal. Gene Ontology for microbial processes useful for bioenergy production. Sixth Annual Department of Energy Joint Genome Institute Users meeting, March 21-24, 2011, Walnut Creek, CA.