Gene Ontology for Microbial Processes Useful for BioEnergy Production

Trudy Torto-Alalibo
Virginia Bioinformatics Institute
What is the Gene Ontology (GO)

The GO provides a controlled universal, uniform language that facilitates high quality functional gene annotation for all species.
The Gene Ontology Consortium (GOC)

Courtesy: GO Consortium
GO is represented by three ontologies

- **Molecular function**: What a gene product does
- **Biological process**: Why it does what it does
- **Cellular component**: where it does what it does
Molecular Function ontology: What the gene product does

- GO:0003674 molecular function [387105 gene products]
- GO:0003824 catalytic activity [150709 gene products]
- GO:0016787 hydrolase activity [53897 gene products]
- GO:0016798 hydrolase activity, acting on glycosyl bonds [6421 gene products]
- GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds [5371 gene products]

▶ GO:0008810 cellulase activity [186 gene products]
Biological Process ontology: Why it does what it does

- GO:0008150 biological_process [364989 gene products]
- GO:0008152 metabolic process [184722 gene products]
- GO:0044238 primary metabolic process [141702 gene products]
  - GO:0005975 carbohydrate metabolic process [19039 gene products]
  - GO:0009056 catabolic process [24808 gene products]
  - GO:0043170 macromolecule metabolic process [102713 gene products]
  - GO:0016052 carbohydrate catabolic process [6055 gene products]
  - GO:0009057 macromolecule catabolic process [8577 gene products]
  - GO:0005976 polysaccharide metabolic process [6437 gene products]
  - GO:0044042 glucan metabolic process [1253 gene products]
  - GO:0000272 polysaccharide catabolic process [2635 gene products]
  - GO:0030243 cellulose metabolic process [327 gene products]
  - GO:0009251 glucan catabolic process [596 gene products]
  - GO:0030245 cellulose catabolic process [264 gene products]
Cellular Component ontology
Where it does what it does

- GO:0005575 cellular_component [332268 gene products]
- GO:0005576 extracellular region [17311 gene products]
  - GO:0044421 extracellular region part [6773 gene products]
  - GO:0043226 organelle [126141 gene products]
    - GO:0043230 extracellular organelle [52 gene products]
    - GO:0043228 non-membrane-bounded organelle [32232 gene products]
      - GO:0043264 extracellular non-membrane-bounded organelle [0 gene products]
      - GO:0043263 cellulosome [0 gene products]
Framework of GO terms

• Each ontology forms the root of carefully defined set of terms
• Each term is related to other terms in a "parent-child" relationship where the child term is more specific than the parent term
• Child terms may have one or more parents - DAG
• A term has a name and detailed definition understandable by humans
• Terms have unique id numbers which makes them easily searchable by computers
General GO tree structure

Root term
  ↓
  general term
  ↓
  less general term
  ↓
  more specific term
  ↓
  even more specific term
  ↓
  most specific term
Sample term representation under the Biological Process ontology

root

1. GO:0008150 biological process [364989 gene products]
   1. GO:0009987 cellular process [198994 gene products]
   1. GO:0008152 metabolic process [184722 gene products]
      1. GO:0044238 primary metabolic process [141702 gene products]
         1. GO:0005975 carbohydrate metabolic process [19039 gene products]
         1. GO:0009056 catabolic process [24808 gene products]
         1. GO:0044237 cellular metabolic process [143271 gene products]
            1. GO:0043170 macromolecule metabolic process [102713 gene products]
               1. GO:0016052 carbohydrate catabolic process [6055 gene products]
               1. GO:0044262 cellular carbohydrate metabolic process [10672 gene products]
               1. GO:0044260 cellular macromolecule metabolic process [88565 gene products]
               1. GO:0009057 macromolecule catabolic process [8577 gene products]
               1. GO:0005976 polysaccharide metabolic process [6437 gene products]
                  1. GO:0044275 cellular carbohydrate catabolic process [3787 gene products]
                  1. GO:0044264 cellular polysaccharide metabolic process [2761 gene products]
                  1. GO:0000272 polysaccharide catabolic process [2635 gene products]

Child terms
General to specific

“true path” rule”: if true for child, must be true for parent
Framework of GO terms

- Each ontology forms the root of carefully defined set of terms
- Each term is related to other terms in a “parent-child” relationship where the child term is more specific than the parent term
- GO defines 3 relationships between terms
- A term has a name and detailed definition understandable by humans
- Terms have unique id numbers which makes them easily searchable by computers
# Term Information

**polysaccharide catabolic process**

<table>
<thead>
<tr>
<th>Term Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accession</strong></td>
</tr>
<tr>
<td><strong>Ontology</strong></td>
</tr>
<tr>
<td><strong>Synonyms</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td><strong>Comment</strong></td>
</tr>
</tbody>
</table>
GO Term Development
GO Term Development

• Can be done by individuals via web based community resource called -Sourceforge

• Undertaken by interest group when major changes have to be made

• As an example PAMGO interest group worked on major changes

• PAMGO established GO node to accommodate terms describing genes in one organism acting in another-microbe-host
PAMGO Approach: Working Groups

- Bacteria-host interactions
- Fungus-host interactions
- Oomycete-host interactions
- Nematode-host interactions

2-weeks

Term proposal discussion via email
Term proposal discussion via teleconferencing
Term proposal discussion via jamboree/meeting

Coordinator

GO Editorial Office
MENGO's MISSION

Generate GO terms and annotate genes associated with microbial processes useful for bioenergy production

Courtesy Biswarup Mukhopadhyay
Suggested MENGO Working Groups

Microbial Biomass Deconstruction

Lignin degradation

Central Metabolism

Molecular Function

Synthetic Biology

Term proposal discussion via email/wiki
Term proposal discussion via teleconferencing
Term proposal discussion via jamboree/meeting

Coordinator

GO Editorial Office

Initiated at 1st MENGO meeting at the Great Lakes Bioenergy Research Center
Term Development-Approach

• Translate major concepts into high level terms
• Pathways/bioconversion processes and enzymatic reactions relevant to these steps
• Check if relevant terms are already in the GO database
• Propose new terms where lacking
Generate GO terms and annotate genes associated with microbial energy processes

Approach
Generate GO terms and annotate genes associated with microbial energy processes

Approach
Biomass deconstruction - cellulose: Level of Granularity?

Biological Process Ontology

- GO:0044042 glucan metabolic process [1215 gene products]
- GO:0000272 polysaccharide catabolic process [2619 gene products]
- GO:0030243 cellulose metabolic process [316 gene products]
- GO:0009251 glucan catabolic process [568 gene products]
- GO:0030245 cellulose catabolic process [252 gene products]

- β-1,4-glycosyl linkages catabolic process
- cellobiose catabolic process
- cellotriose catabolic process
- regulation of cellulose catabolic process
- positive regulation of cellulose catabolic process
- negative regulation of cellulose catabolic process
Generate GO terms and annotate genes associated with microbial energy processes

Approach
Hemicellulose Hydrolysis

Biological Process Ontology

- GO:0005976 polysaccharide metabolic process [6428 gene products]
- GO:0010383 cell wall polysaccharide metabolic process [514 gene products]
- GO:0010410 hemicellulose metabolic process [443 gene products]
  - GO:0010412 mannan metabolic process [123 gene products]
  - GO:0045491 xylan metabolic process [311 gene products]
  - GO:0010411 xyloglucan metabolic process [10 gene products]

hemicellulose catabolic process
mannan catabolic process
xylan catabolic process
xyloglucan catabolic process
Hemicellulase Activity

Molecular Function ontology

Hemicellulase activity
endo-1,4-β-d-xylanase activity
exo-1,4-β-d-xylanase activity
1,4-β-d-xylosidase activity
1,4-β-d-mannosidase activity
Term Development - Example

Generate GO terms and annotate genes associated with microbial energy processes

Approach

Courtesy Biswarup Mukhopadhyay
Fermentative Pathway

Identify specific steps and components for BP and/or MF terms
Missing Synonyms

• Include synonyms that represent Bioenergy processes eg.
  • Saccharification
  • Deconstruction
  • Solventogenesis
Suggested MENG0 Working Groups

Microbial Biomass Deconstruction

Lignin degradation

Central Metabolism

Molecular Function

Synthetic Biology

Protein Structure

Term proposal discussion via email/wiki
Term proposal discussion via teleconferencing
Term proposal discussion via jamboree/meeting

Coordinator

GO Editorial Office

Initiated at 1st MENG0 meeting at the Great Lakes Bioenergy Research Center
http://wiki.geneontology.org/index.php/Category:Content_MENG0
Annotation of Gene Products with the Gene Ontology
There is a lot of research on microbial processes useful for bioenergy.
To Make a GO Annotation

• Achieve a comprehensive picture of the available information for a given gene product

• framed within a controlled vocabulary and ready for comparison with other annotated gene products

• Can use a more general or more specific terms in a tree, depending on level of information available
## Elements of Annotation

<table>
<thead>
<tr>
<th>GO ID</th>
<th>Ontology type</th>
<th>Term</th>
<th>Evidence code</th>
<th>Reference</th>
<th>With/from</th>
<th>Taxon/Dual taxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:####</td>
<td>MF</td>
<td>Full name of term</td>
<td>Experimental</td>
<td>PMID:#####</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td></td>
<td>evidence codes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td></td>
<td>IDA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IPI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IGI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IMP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IEP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ISS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IEA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Captures additional info including:

- Interacting proteins (IGI, IPI)
- Similar proteins (ISS)

*These field is not used for all annotations*

GO Evidence Codes


Experimental Evidence Codes:
Inferred from Experiment [EXP]
Inferred from Direct Assay [IDA]
Inferred from Physical Interaction [IPI]
Inferred from Mutant Phenotype [IMP]
Inferred from Genetic Interaction [IGI]
Inferred from Expression Pattern [IEP]

Computational Analysis Evidence Codes:
Inferred from Sequence or Structural Similarity [ISS]
Inferred from Sequence Alignment [ISA]
Inferred from sequence orthology [ISO]
Inferred from Genomic Context [IGC]
Inferred from Reviewed Computational Analysis [RCA]
(Inferred from Electronic Annotation [IEA] - no curatorial review)

Author Statement Evidence Codes
Traceable [TAS] or Non-traceable Author Statement [NAS]

Curatorial Statement Evidence Codes
Inferred by Curator [IC] or No Biological Data Available [ND]
Novel bacterial ferulic acid esterase from *Cellvibrio japonicus* and its application in ferulic acid release and xylan hydrolysis

Shara D. McClendon · Hyun-Dong Shin · Rachel R. Chen

Received: 8 June 2010 / Accepted: 24 August 2010 / Published online: 7 September 2010
© Springer Science+Business Media B.V. 2010
Ferulic acid esterase (FAE) activity was determined by measuring the amount of ferulic acid released from methyl ferulate (Apin Chemicals, Oxon, UK). Reactions were carried out by adding enzyme to 100 mM sodium phosphate buffer, pH 7.0, with 1 mM methyl ferulate and incubating the mixture at 40°C for 30 min. The reaction was stopped by adding glacial acetic acid. FAE activity on complex substrates was performed by adding 20 mU/mg enzyme to 10 mg wheat arabinoxylan (WAX)/ml or (destarched) corn bran (CB, DCB) and incubating at 40°C with and without presence of xylanase (0.1 U/mg substrate) in a thermomixer for 24 h at 1400 rpm. The reaction was ended by boiling the mixture for 5 min and collecting the supernatant via centrifuga-
Experimental Data Supporting Function of Fee1B

Methods. Fee1B was successfully expressed in E. coli BL21(DE3) as a His-tagged fusion protein and purified using immobilized metal affinity chromatography. The one-step affinity chromatography yielded a purified protein with an expected molecular weight of 61 kDa and an overall recovery of 60% (Table 1). The purified recombinant enzyme was active toward a common FAE synthetic substrate, methyl ferulate, with about 28 mU/mg protein, indicating that the gene fee1B encodes a functional FAE.

Table 2: Substrate specificity of Fee1B from C. japonicus

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Specific activity&lt;sup&gt;a&lt;/sup&gt; (mU/mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fee1B</td>
</tr>
<tr>
<td>p-Nitrophenyl acetate</td>
<td>50 ± 2</td>
</tr>
<tr>
<td>Methyl ferulate</td>
<td>27.5 ± 1.7</td>
</tr>
<tr>
<td>Methyl p-coumarate</td>
<td>12.4 ± 0.8</td>
</tr>
<tr>
<td>Methyl caffeate</td>
<td>5.4 ± 0.3</td>
</tr>
<tr>
<td>Methyl sinapate</td>
<td>10.5 ± 1.9</td>
</tr>
</tbody>
</table>

<sup>a</sup> Each reaction was performed in duplicate

Ferulic acid esterase substrates
Feruloyl esterase activity

**Term Information**

<table>
<thead>
<tr>
<th>Accession</th>
<th>GO:0030600</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology</td>
<td>Molecular Function</td>
</tr>
</tbody>
</table>
| Synonyms   | exact: 4-hydroxy-3-methoxycinnamoyl-sugar hydrolase activity  
related: cinnAE  
exact: cinnamoyl ester hydrolase activity  
related: FAE-I  
related: FAE-II  
related: FAE-III  
related: FAEA  
exact: ferulic acid esterase activity  
related: hemicellulase accessory  
exact: hydroxycinnamoyl esterase activity |
| Definition  | Catalysis of the reaction: feruloyl-polysaccharide + H2O = ferulate + polysaccharide.  
Source: EC:3.1.1.73 |
## Annotation of FeelB

<table>
<thead>
<tr>
<th>GO ID</th>
<th>Ontology type</th>
<th>Term</th>
<th>Evidence code</th>
<th>Reference</th>
<th>With/from</th>
<th>Taxon/Dual taxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0030600</td>
<td>M</td>
<td>Feruloyl esterase activity</td>
<td>IDA</td>
<td>PMID: 20821249</td>
<td></td>
<td>155077</td>
</tr>
</tbody>
</table>

Unique information about gene product should be described in different rows
GENE ONTOLOGY TOOLS
Gene Ontology Tools


Consortium Tools
- Tools for searching and browsing GO
- Annotation tools
- Tools for gene expression/microarray analysis
- Other tools

Non-Consortium Tools
- Tools for searching and browsing GO
- Annotation tools
- Tools for gene expression/microarray analysis
- Other tools

Key

Windows, X
AmiGO provides an interface to search and browse the ontology and annotation data provided by the GO consortium. Users can search for gene products and view the terms with which they are associated; alternatively, users can search or browse the ontology for GO terms of interest and see term details and gene product annotations. AmiGO also provides a BLAST search engine, which
Search the Gene Ontology database

Endo-1,4-β-xylanase activity

- GO terms
- genes or proteins
- exact match

Submit Query
AMIGO: The GO Browser
http://amigo.geneontology.org

Inferred Tree View

Mappings

1. GO:0003674 molecular_function [376402 gene products]
   1. GO:0003824 catalytic activity [148649 gene products]
   1. GO:0016787 hydrolase activity [53071 gene products]
   1. GO:0016798 hydrolase activity, acting on glycosyl bonds [6292 gene products]
   1. GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds [5271 gene products]

1. GO:0031176 endo-1,4-beta-xylanase activity [163 gene products]
AMIGO: The GO Browser
http://amigo.geneontology.org

endo-1,4-beta-xylanase activity

Gene Product Associations to endo-1,4-beta-xylanase activity; GO:0031176 and children

Download all association information in: gene association format, RDF-XML

Filter associations displayed
Filter by Gene Product
Gene Product Type
All
gene
gene product
mirna
Data source
All
ASAP
AspGD
CGD
Species
All
Arabidopsis thaliana
Bacillus anthracis
Bacillus subtilis
Filter by Association
Evidence Code
All
IC
IDA
IEA
View associations
All
Direct associations
Set filters
Remove all filters
## AMIGO: The GO Browser

http://amigo.geneontology.org

### endo-1,4-beta-xylanase activity; GO:0031176

<table>
<thead>
<tr>
<th>Symbol, full name</th>
<th>Information</th>
<th>Qualifier</th>
<th>Evidence</th>
<th>Reference</th>
<th>Assigned by</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 associations</td>
<td>protein from Aspergillus niger</td>
<td>ISS With UniProtKB:O94163</td>
<td>GO REF:0000024</td>
<td>UniProtKB</td>
<td></td>
</tr>
<tr>
<td>xlnC</td>
<td>Probable endo-1,4-beta-xylanase C</td>
<td>ISS With UniProtKB:O94163</td>
<td>GO REF:0000024</td>
<td>UniProtKB</td>
<td></td>
</tr>
<tr>
<td>3 associations</td>
<td>protein from Aspergillus terreus NIH2624</td>
<td>ISS</td>
<td>GO REF:0000024</td>
<td>UniProtKB</td>
<td></td>
</tr>
<tr>
<td>xlnC</td>
<td>Probable endo-1,4-beta-xylanase C</td>
<td>ISS With UniProtKB:O94163</td>
<td>GO REF:0000024</td>
<td>UniProtKB</td>
<td></td>
</tr>
<tr>
<td>3 associations</td>
<td>gene_product from Emericella nidulans</td>
<td>RCA</td>
<td>PMID:18405346</td>
<td>ASPGD (via AspGD)</td>
<td></td>
</tr>
<tr>
<td>xlnE</td>
<td>Probable endo-1,4-beta-xylanase C</td>
<td>ISS With UniProtKB:O94163</td>
<td>GO REF:0000024</td>
<td>UniProtKB</td>
<td></td>
</tr>
<tr>
<td>3 associations</td>
<td>protein from Magnaporthe grisea</td>
<td>IMP</td>
<td>PMID:16461639</td>
<td>PAMGO MGG (via NCBI)</td>
<td></td>
</tr>
<tr>
<td>XYL-6</td>
<td>Probable endo-1,4-beta-xylanase C</td>
<td>ISS With UniProtKB:O94163</td>
<td>GO REF:0000024</td>
<td>UniProtKB</td>
<td></td>
</tr>
<tr>
<td>2 associations</td>
<td>protein from Aspergillus niger</td>
<td>IDA</td>
<td>DOI:10.1023/A:1018327623422</td>
<td>UniProtKB</td>
<td></td>
</tr>
</tbody>
</table>
GO Tools for Analyses of Omics Data

- Exploratory gene association networks
- Expression Profiler
- FuncExpression
- FunCluster
- Functional Analysis of Transcriptional Networks
- Gene Class Expression
- Gene Ontology Enrichment Analysis Software Toolkit
MENGO Interest Group: 
Get Plugged In

MENGO Workshops

• February 10, 2011: Great Lakes Bioenergy Research Center
• March 25th, 2011: Joint Bioenergy Institute as part of the JGI users meeting
• April 13–14, 2011 as part of the DOE PI meeting in Washington D.C

Join the term development discussion Subscribe to mengo-discussion@vbi.vt.edu

Lead/Join a working group:
Acknowledgements

**MENGO Team**
- Joao Setubal
- Biswarup Mukhopadhyay
- Brett Tyler
- T. M. Murali
- Trudy Torto-Alalibo
- Endang Purwantini

**Collaborators**
- Caroline Hardwood (Univ. of Washington)
- Robert Blankenship (Washington University)
- Scientific Community working on BIOENERGY processes

**Funding**
- Office of Science (BER), U.S. Department of Energy
BRAINSTORM SESSION