

Predicted Phenotypes

Part of the content of this user guide was taken from the following IMG paper. For more information on phenotype rules and prediction, please refer to: [PLoS ONE 8\(2\): e54859](https://doi.org/10.1371/journal.pone.0054859).
[doi:10.1371/journal.pone.0054859](https://doi.org/10.1371/journal.pone.0054859)

Pathways can be best understood in context of other pathways within the organism. For example, if an organism degrades cellulose to cellobiose outside the cell, it can only utilize cellulose as carbon source if it also has a transport pathway for uptake of cellobiose and, within the cell, a metabolic pathway to gain energy from cellobiose. If all steps are present, then the organism will have the phenotype of cellulose utilization with cellobiose as an intermediate. Phenotypes correspond directly to biological traits that can be measured in an experiment, and thus provide a powerful mechanism for further assessing the coherence of functional and pathways annotations.

IMG pathways provide the context needed for predicting phenotypes within the IMG system. For example, for the transport step in the cellulose utilization via cellobiose there are two known possibilities: cellobiose can be taken up by an ABC transporter or by the phosphotransferase system. Within the cell, cellobiose can either be converted to glucose and glucose 6-phosphate, glucose and glucose 1-phosphate, or two molecules of glucose. So the phenotype of cellulose utilization can be specified in IMG using the following set of logical rules reflecting different combination of pathways, as illustrated in Figure 1(i):

Phenotype Rule (i)

Growth on cellulose via cellobiose

Rule ID	00051
Name	Growth on cellulose via cellobiose
Category	Metabolism
Category Value	Cellobiose degrader
Description	Organism is predicted to be able to use cellulose as carbon and energy source with cellobiose as an intermediate
Add Date	2011-03-01
Last Mod Date	2011-03-01
Modified By	Iain Anderson (IAnderson@lbl.gov)

Rule

([IPWAY:567](#): Cellulose degradation to cellobiose)
and [IPWAY:732](#): ATP-dependent cellobiose uptake
and [IPWAY:554](#): Cellobiose hydrolysis)
OR ([IPWAY:567](#): Cellulose degradation to cellobiose
and [IPWAY:732](#): ATP-dependent cellobiose uptake
and [IPWAY:553](#): Cellobiose conversion to glucose and glucose 1-phosphate)
OR ([IPWAY:567](#): Cellulose degradation to cellobiose
and [IPWAY:732](#): ATP-dependent cellobiose uptake
and [IPWAY:909](#): Cellobiose phosphorylation via beta-glucoside kinase
and [IPWAY:552](#): Cellobiose 6-phosphate conversion to glucose and glucose 6-phosphate)
OR ([IPWAY:567](#): Cellulose degradation to cellobiose
and [IPWAY:669](#): Cellobiose uptake via phosphotransferase system
and [IPWAY:552](#): Cellobiose 6-phosphate conversion to glucose and glucose 6-phosphate)

Phenotype Rule (ii)

L-histidine prototroph

Rule ID	00001
Name	L-histidine prototroph
Category	Metabolism
Category Value	Prototrophic
Description	Organism is predicted to be able to synthesize L-histidine.
Add Date	2009-04-08
Last Mod Date	2009-11-20
Modified By	Iain Anderson (IAnderson@lbl.gov)

Rule

([IPWAY:162](#): L-histidine synthesis)

Figure 1. Phenotype Rules

1. Cellulose degradation to cellobiose **AND** ATP-dependent cellobiose uptake **AND** Cellobiose hydrolysis; **or**
2. Cellulose degradation to cellobiose **AND** ATP-dependent cellobiose uptake **AND** Cellobiose conversion to glucose and glucose 1-phosphate; **or**
3. Cellulose degradation to cellobiose **AND** ATP-dependent cellobiose uptake **AND** Cellobiose phosphorylation via beta-glucoside kinase **AND** Cellobiose 6-phosphate conversion to glucose and glucose 6-phosphate; **or**
4. Cellulose degradation to cellobiose **AND** Cellobiose uptake via phosphotransferase system **AND** Cellobiose 6-phosphate conversion to glucose and glucose 6-phosphate.

Following this approach we have developed a system for recording IMG phenotype definitions and inferring organism phenotypes based on these rules and pathway assertion status. In general, IMG phenotype definitions consist of IMG pathway status values connected by logical operators **AND**, **OR** and **NOT**. The simplest IMG phenotype definition consists of one IMG pathway status value without operators. Consider the example of histidine biosynthesis, for which only one pathway is known. In this case the phenotype of L-histidine prototrophy (i.e., organism's ability to synthesize histidine) can be recorded simply as the presence of the corresponding pathway (IMG pathway 162), as illustrated in Figure 1(ii). In contrast, the phenotype of L-histidine auxotrophy (i.e., organism's inability to synthesize

histidine) is recorded as NOT (IMG pathway 162). Given the 3-value status of IMG pathways (“asserted”, “unknown”, “not asserted”), evaluation of the statement **NOT** (IMG pathway 162) is performed as shown below:

NOT	
IMG Pathway Assertion	Evaluation Result
Asserted	False
Not asserted	True
Unknown	Unknown

Based on this evaluation an organism will be assigned the IMG phenotype of L-histidine auxotroph if IMG pathway 162 has the status “not asserted”, whereas an organism in which the status of IMG pathway 162 is “asserted” will be assigned the IMG phenotype L-histidine prototroph. Both phenotype predictions can be readily tested by attempting to grow an organism on a medium with and without L-histidine: while L-histidine prototrophs will successfully grow on both media, L-histidine auxotrophs should fail to grow in the absence of externally provided histidine. For IMG phenotypes consisting of two pathway status values connected by operators **OR** or **AND**, evaluations are performed as shown below:

OR			
Evaluation Result	Pathway 2: Asserted	Pathway 2: Not Asserted	Pathway 2: Unknown
Pathway 1: Asserted	True	True	True
Pathway 1: Not Asserted	True	False	Unknown
Pathway 1: Unknown	True	Unknown	Unknown

AND			
Evaluation Result	Pathway 2: Asserted	Pathway 2: Not Asserted	Pathway 2: Unknown
Pathway 1: Asserted	True	False	Unknown
Pathway 1: Not Asserted	False	False	False
Pathway 1: Unknown	Unknown	False	Unknown

More complex phenotypes can be recorded using multiple pathway status values connected by multiple AND, OR and NOT operators, whereby the result of each operation is evaluated as described above. For example, in order to be a phenylalanine prototroph an organism must synthesize chorismate and then synthesize phenylalanine from chorismate. For both of these steps, there are two possible pathways, so the phenotype L-phenylalanine prototroph is an AND-rule that can be recorded using IMG pathway identifiers as: (IMG pathway 146 OR IMG pathway 519) AND (IMG pathway 272 OR IMG pathway 147). Another example of a complex phenotype is L-lysine prototrophy, which requires the presence of at least 1 out of 6 possible biosynthetic pathways, and can be recorded as (IMG pathway 169 OR IMG pathway 170 OR IMG pathway 0171 OR IMG pathway 199 OR IMG pathway 333 OR IMG pathway 465). On the other hand, the phenotype of L-lysine auxotrophy is recorded as (NOT IMG pathway 169) AND (NOT IMG pathway 170) AND (NOT IMG pathway 0171) AND (NOT IMG pathway 199) AND (NOT IMG pathway 333) AND (NOT IMG pathway 465).

From the **Find Functions** menu, users can click on the **Phenotypes** option to view all predicted phenotypes in IMG as defined by JGI scientists (see Figure 2).

Predicted Phenotypes  

A phenotype is broadly defined as an observable characteristic of an organism. Current phenotypes in IMG are predicted using a set of rules based on IMG's collection of pathways and parts. Follow the link provided by **Rule ID** to view the information for the selected rule. Follow the link provided by **No. of Genomes w/ Phenotype** to the list of genomes associated with a specific phenotype. Click on the number to view the results in tree display, or click on  to view the results in table display.

For more information on phenotype rules and prediction, please refer to: [PLoS ONE 8\(2\): e54859. doi:10.1371/journal.pone.0054859](https://doi.org/10.1371/journal.pone.0054859)

Filter column: Rule ID Filter text Apply 

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector

Rule ID	Name	Category	Category Value	Description	No. of Genomes w/ Phenotype
00001	L-histidine prototroph	Metabolism	Prototrophic	Organism is predicted to be able to synthesize L-histidine.	6354 
00002	Aerobe	Oxygen Requirement	Aerobe	Organism is predicted to be able to grow in the presence of air.	33318 
00003	L-lysine prototroph	Metabolism	Prototrophic	Organism is predicted to be able to synthesize L-lysine.	8623 
00004	Denitrifier	Metabolism	Denitrifying	Organism is predicted to be able to reduce nitrate to nitrogen (N2).	1347 
00005	Use of nitrate as electron acceptor	Metabolism	Nitrate reducer	Organism is predicted to be able to grow anaerobically with nitrate as electron acceptor	17360 
00006	Carbon fixation	Metabolism	Carbon fixation	Organism is predicted to be able to use carbon dioxide as sole carbon source	327 
00007	L-lysine auxotroph	Metabolism	Auxotroph	Organism is predicted to be unable to synthesize L-lysine.	46566 
00010	L-alanine prototroph	Metabolism	Prototrophic	Organism is predicted to be able to synthesize	34076 

Figure 2. Phenotype List

There are 6 columns in the list:

- **Rule ID:** Phenotype rule ID. Clicking on the ID will lead to Phenotype Rule definition page as shown in Figure 1(i) and Figure 1(ii).
- **Name:** Name of the phenotype rule.
- **Category:** Rule category such as Metabolism or Oxygen Requirement.
- **Category Value:** Rule category value. For example, for Rule 10 *L-alanine prototroph* is Prototrophic, while its opposite Rule 11 *L-alanine auxotroph* is Auxotroph.
- **Description:** Description of this rule.
- **No. of Genomes w/ Phenotypes:** Number of isolate genomes that have been predicted to have the phenotype. IMG users can view all genomes with predicted phenotype in a table or phylogenetic tree hierarchical display. Click on the number to view in tree display (Figure 3(i)), and click on the table icon to view in table display (Figure 3(ii)).

(i)

Genomes with Selected Phenotype

Phenotype Rule: [L-histidine prototroph \(00001\)](#)

hint: Click on (v) or (x) to see rule prediction detail; (v) indicates that the genome has the phenotype, and (x) indicates that the genome does not have the phenotype. [Display genomes with selected phenotype in complete tree.](#)

Domains(D): B = Bacteria, A = Archaea, E = Eukarya.
Genome Completion: [F]inished, [P]ermanent Draft, [D]raft.

- 01 **Archaea**
 - 02 Euryarchaeota
 - 03 Halobacteria
 - 04 Haloferacales
 - 05 Halorubraceae
 - 06 Halorubrum
 - 08 [Halorubrum lacusprofundi ATCC 49239 \(F\) \(JGI\) -- L-histidine prototroph \(v\)](#)
 - 04 Methanobacteria
 - 04 Methanobacteriales
 - 05 Methanobacteriaceae
 - 06 Methanobrevibacter
 - 08 [Methanobrevibacter sp. 87.7 \(Genome assembly of Methanobrevibacter sp. 87.7\) \(D\) -- L-histidine prototroph \(v\)](#)
- 01 **Bacteria**
 - 02 Actinobacteria
 - 04 Actinomycetales
 - 05 Actinomycetaceae
 - 06 Actinomycetes
 - 08 [Actinomycetes naeskundi MG1 \(Chromosome\) \(P\) -- L-histidine prototroph \(v\)](#)
 - 08 [Actinomycetes ruminicola DSM 27982 \(P\) \(JGI\) -- L-histidine prototroph \(v\)](#)

(ii)

Genomes with Selected Phenotype

Phenotype Rule: [L-histidine prototroph \(00001\)](#)

hint: Follow the link provided by **Phenotype** to see rule prediction details. [Display genomes with selected phenotype in complete tree.](#)

Domains(D): B = Bacteria, A = Archaea, E = Eukarya.

6354 of 6354 rows selected

Filter column: Genome ID Filter text Apply

Export Page 1 of 64 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	Domain	Genome ID	Genome Name	Last Modified On	Phenotype
<input checked="" type="checkbox"/>	B	637000007	Anaeromyxobacter dehalogenans 2CP-C	2015-10-05	L-histidine prototroph
<input checked="" type="checkbox"/>	B	637000019	Bacillus clausii KSM-K16	2015-10-05	L-histidine prototroph
<input checked="" type="checkbox"/>	B	637000020	Bacillus halodurans C-125	2015-10-05	L-histidine prototroph
<input checked="" type="checkbox"/>	B	637000024	Bacteroides fragilis NCTC 9343	2015-10-05	L-histidine prototroph
<input checked="" type="checkbox"/>	B	637000025	Bacteroides fragilis YCH46	2015-10-05	L-histidine prototroph
<input checked="" type="checkbox"/>	B	637000026	Bacteroides thetaiotaomicron VPI-5482	2015-10-05	L-histidine prototroph
<input checked="" type="checkbox"/>	B	637000043	Buchnera aphidicola APS	2015-10-05	L-histidine prototroph

Figure 3. Genomes Predicted with Phenotype

Viewing phenotype prediction using a hierarchical tree display has the advantage of showing the distribution of phenotypes in genetically related organisms. There are two hierarchical display options. The first display option only shows organisms with the predicted phenotype, which is more compact and easier to view, as illustrated in Figure 3(i). The second display option shows all organisms (select the complete tree display link in **Hint** for this option); genomes with predicted phenotypes are selected and shown with the phenotype label. This hierarchical display allows users to compare closely related organisms with and without the predicted phenotype. In both display modes, users can click on the mark (v) to see phenotype prediction evaluation.

Users can select a subset of genomes to add to Genome Cart or Workspace for further analysis.