

## 1) Searching for TroutBog Metagenomes in IMG/MER

2) Click on one metagenome

### 3) Click on Protein coding genes → with product names:

	Number	% of Unassembled
Number of sequences	578764	100.00%
Number of bases	98661387	100.00%
GC count	47873305	48.52%
<b>Genes</b>		
RNA genes	2508	0.42%
rRNA genes	1303	0.22%
5S rRNA	147	0.02%
16S rRNA	387	0.07%
18S rRNA	14	0.00%
23S rRNA	718	0.12%
28S rRNA	37	0.01%
tRNA genes	1205	0.20%
Protein coding genes	<a href="#">592481</a>	99.58%
with Product Name	<a href="#">175791</a>	29.55%
with COG	<a href="#">174846</a>	29.39%
with Pfam	<a href="#">231560</a>	38.92%
with KO	<a href="#">168851</a>	28.38%
with Enzyme	<a href="#">104640</a>	17.59%
with MetaCyc	<a href="#">68836</a>	11.57%
with KEGG	<a href="#">106213</a>	17.85%
COG Clusters	3998	86.33%
Pfam Clusters	12245	75.10%

#### Search Protein Coding Genes

 **hint:** Genes with hypothetical protein are not included in the gene product name search result.

Data Type	unassembled ▾
Keyword	
Filter	Gene Product Name (inexact) *

Include the following information in the result:

Gene Detailed Information

[https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=MetaDetail&page=withFunc&taxon\\_oid=3300000643&data\\_type=unassembled&total\\_genome\\_gene\\_count=594989&total\\_gene\\_count=175791](https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=MetaDetail&page=withFunc&taxon_oid=3300000643&data_type=unassembled&total_genome_gene_count=594989&total_gene_count=175791)

# 4) Searching for all nitrogenase genes present in the chosen Trout Bog metagenome

The screenshot shows the JGI/IMG/MER web interface. At the top, there is a navigation bar with links for Home, Find Genomes, Find Genes, Find Functions, Compare Genomes, OMICS, Workspace, My IMG, Data Marts, and Help. A user is logged in as Leyden Fernandez. A search bar at the top right allows for quick genome searching.

The main content area displays a table titled "Genes with Function Prediction (unassembled)" for the genome "Freshwater microbial communities from Trout Bog Lake, WI - 05NOV2007 hypolimnion (MER-FS) (unassembled)".

Below the title, there are buttons for "Add Selected to Gene Cart", "Select All", and "Clear All". A filter is applied to the "Gene ID" column, with the search term "Nitrogenase" highlighted in red. The results table shows 10 rows of gene predictions, each with a checkbox for selection and a link to the gene's product name.

Select	Gene ID	Gene Product Name
<input type="checkbox"/>	<a href="#">B570J11626_30000471</a>	AcrB/AcrD/AcrF family
<input type="checkbox"/>	<a href="#">B570J11626_30000841</a>	ABC-2 type transporter
<input type="checkbox"/>	<a href="#">B570J11626_30001121</a>	Uncharacterized protein conserved in bacteria
<input type="checkbox"/>	<a href="#">B570J11626_30001421</a>	tRNA_Arg_CCT
<input type="checkbox"/>	<a href="#">B570J11626_30001451</a>	Hydrogenase maturation factor
<input type="checkbox"/>	<a href="#">B570J11626_30001501</a>	Predicted N6-adenine-specific DNA methylase
<input type="checkbox"/>	<a href="#">B570J11626_30001601</a>	3-oxoacyl-(acyl-carrier-protein) synthase
<input type="checkbox"/>	<a href="#">B570J11626_30002071</a>	Tetraacyldisaccharide-1-P 4'-kinase
<input type="checkbox"/>	<a href="#">B570J11626_30003741</a>	ATPase related to the helicase subunit of the Holliday junction resolvase
<input type="checkbox"/>	<a href="#">B570J11626_30003751</a>	Tryptophanyl-tRNA synthetase
<input type="checkbox"/>	<a href="#">B570J11626_30003761</a>	FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor)

Download the nitrogenase genes annotation (GeneCart\* is call the file after downloading from IMG)

The screenshot shows two windows of the GeneCart software. The left window is a search results table with columns: Gene ID, Gene Product Name, and Genome ID. It lists numerous entries for 'Nitrogenase' genes across different genomes. The right window is a detailed view of a specific gene entry, showing a table with columns: Gene ID, Locus Tag, Gene Product Name, and Genome ID. This view includes a 'Find' bar at the top and a scrollable list of annotations below.

Gene ID	Gene Product Name	Genome ID
B570J11626_30468141	Dinitrogenase iron-molybdenum cofactor	33000007
B570J11626_30654361	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_30741161	Nitrogenase component 1 type Oxidoreductase	33000007
B570J11626_31494351	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_31524911	Nitrogenase component 1 type Oxidoreductase	33000007
B570J11626_31578161	Dinitrogenase reductase ADP-ribosyltransferase (DRAT)	33000007
B570J11626_32022781	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_32032441	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_32135321	Nitrogenase component 1 type Oxidoreductase	33000007
B570J11626_32586811	Nitrogenase subunit NiFH (ATPase)	33000007
B570J11626_33381121	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_33421191	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_33556271	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_33963951	Nitrogenase component 1 type Oxidoreductase	33000007
B570J11626_34048531	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_34619941	Dinitrogenase iron-molybdenum cofactor	33000007
B570J11626_34991971	Nitrogenase component 1 type Oxidoreductase	33000007
B570J11626_35738611	Mo-dependent nitrogenase C-terminus	33000007
B570J11626_35838801	Nitrogenase component 1 type Oxidoreductase	33000007
B570J11626_35945461	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007
B570J11626_36141761	Nitrogenase molybdenum-iron protein, alpha and beta chains	33000007

Download the excel file with the annotations here

Example of the nitrogenase Annotations (GeneCart\* files)

## Gene Information

Gene Information	
Gene ID	B570J11626_30468141
Gene Symbol	
Locus Tag	B570J11626_30468141
Product Name	Dinitrogenase iron-molybdenum cofactor
IMG Product Source	pfam02579
Assembled/Unassembled	unassembled
Genome	<a href="#">Freshwater microbial communities from Trout Bog Lake, WI - 05NOV2007 hypolimnion</a>
DNA Coordinates	1..81 (+)(81bp)
IMG ORF Type	
GC Content	0.62
Estimated Copy 	1
Protein Information	
Amino Acid Sequence Length	27aa
Pfam	<a href="#">pfam02579 - Nitro_FeMo-Co</a>
Pathway Information	
KEGG Orthology (KO) Term	<a href="#">KO:K02585 nifB nitrogen fixation protein NifB</a>
KEGG Pathway	
KEGG Orthology (KO) Modules	

[Add To Gene Cart](#)

## Find Candidate Product Name

Method:

Display Option (for sequence based only):

[Find Candidate Product Name](#)

## Evidence For Function Prediction

Click on Find Candidate.... To get the BLAST results which contain taxonomy, percentage of identity, annotations: GO, Pfam and KO

## Candidate Product Names for Query Gene (OID: B570J11626\_30468141)

Genome Name	Freshwater microbial communities from Trout Bog Lake, WI - 05NOV2007 hypolimnion
Gene Product Name	Dinitrogenase iron-molybdenum cofactor
TIGRFam	(pfam02579) Dinitrogenase iron-molybdenum cofactor
KEGG Orthology (KO)	nitrogen fixation protein NifB

Domains(D): \* = Microbiome,  
 B = Bacteria, A = Archaea, E = Eukarya, P = Plasmids, G = GFragment, V = Viruses.  
 Genome Completion(C): F = Finished, P = Permanent Draft, D = Draft.

Sort column: Bit Score ▾ Filter text ▾ Apply ?

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

Column Selector Select Page Deselect Page

Select	Homolog Gene	Homolog Product Name	IMG Term OID	IMG Term	Domain	Status	Genome	Percent Identity	Alignment On Query Gene	Alignment On Homolog Gene	E-value	Bit Score	TIGRFam	COG	Pfam
<input checked="" type="radio"/>	<a href="#">2582993927</a>	Predicted Fe-Mo cluster-binding protein, NifX family			B	D	<a href="#">Composite genome from Trout Bog Hypolimnion pan-assembly TBhypo.metabat.433</a>	100.00			2.20e-13	66		Predicted Fe-Mo cluster-binding protein, NifX family	Dinitrogenase iron-molybdenum cofactor
<input checked="" type="radio"/>	<a href="#">649851827</a>	nitrogenase cofactor biosynthesis protein NifB			B	F	<a href="#">Desulfovibrio aestuoleensis Aspo-2</a>	76.92			3.50e-08	51			Dinitrogenase iron-molybdenum cofactor Radical SAM superfamily
<input checked="" type="radio"/>	<a href="#">2563383501</a>	nitrogen fixation protein NifB			B	P	<a href="#">Desulfonatronovibrio hydrogenovorans DSM 9292</a>	73.08			3.50e-08	51		Radical SAM superfamily enzyme, MoaA/NifB/PqqE/SkfB family	Dinitrogenase iron-molybdenum cofactor Radical SAM superfamily

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



Select the genes and Click on Export the file. It will download a file with the name missingGenes\*. See examples in the folder of GitHub IMG\_analysis and in the next slide

# Example of the missing\* files:

Screenshot of LibreOffice Calc showing a list of missing gene files and their details.

The window title is "missingGenes94614\_09-mar-2017.xls - LibreOffice Calc". The status bar shows the date and time: "mån 18 feb 2019 20.49.45".

The spreadsheet has two tabs: "Sheet 1 of 1" and "missingGenes94614\_09-mar-2017.xls".

**Sheet 1 of 1:**

Name	Size	Type	Modified
GeneCart93972_09-mar-2017.xls	58,3 kB	Spreadsheet	mar 9 2017
genelist93893_09-mar-2017.xls	17,2 kB	Spreadsheet	mar 9 2017
IHXY_potential_NIFH	69,2 kB	Text	maj 18 2017
IHXY_potential_NIFH.output	11,6 kB	Text	maj 19 2017
missingGenes94614_09-mar-2017.xls	11,2 kB	Spreadsheet	mar 9 2017
missingGenes94641_09-mar-2017.xls	11,8 kB	Spreadsheet	mar 9 2017
missingGenes94671_09-mar-2017.xls	11,6 kB	Spreadsheet	mar 9 2017
missingGenes94676_09-mar-2017.xls	1,4 kB	Spreadsheet	mar 9 2017
missingGenes94688_09-mar-2017.xls	10,5 kB	Spreadsheet	mar 9 2017
missingGenes94814_09-mar-2017.xls	12,1 kB	Spreadsheet	mar 9 2017
missingGenes94841_09-mar-2017.xls	12,4 kB	Spreadsheet	mar 9 2017
missingGenes94880_09-mar-2017.xls	11,2 kB	Spreadsheet	mar 9 2017
missingGenes94898_09-mar-2017.xls	15,0 kB	Spreadsheet	mar 9 2017
missingGenes94929_09-mar-2017.xls	11,0 kB	Spreadsheet	mar 9 2017
missingGenes95066_09-mar-2017.xls	57,7 kB	Spreadsheet	mar 9 2017
missingGenes95096_09-mar-2017.xls	13,1 kB	Spreadsheet	mar 9 2017
missingGenes95144_09-mar-2017.xls	13,3 kB	Spreadsheet	mar 9 2017
missingGenes95156_09-mar-2017.xls	11,7 kB	Spreadsheet	mar 9 2017
missingGenes95165_09-mar-2017.xls	15,5 kB	Spreadsheet	mar 9 2017
missingGenes95182_09-mar-2017.xls	12,6 kB	Spreadsheet	mar 9 2017
missingGenes95225_09-mar-2017.xls	10,4 kB	Spreadsheet	mar 9 2017
missingGenes95251_09-mar-2017.xls	11,9 kB	Spreadsheet	mar 9 2017
missingGenes95277_09-mar-2017.xls	11,5 kB	Spreadsheet	mar 9 2017
missingGenes95398_09-mar-2017.xls	11,0 kB	Spreadsheet	mar 9 2017
missingGenes95842_09-mar-2017.xls	11,0 kB	Spreadsheet	mar 9 2017
missingGenes95878_09-mar-2017.xls	11,8 kB	Spreadsheet	mar 9 2017
missingGenes95912_09-mar-2017.xls	11,6 kB	Spreadsheet	mar 9 2017
missingGenes95935_09-mar-2017.xls	12,8 kB	Spreadsheet	mar 9 2017
missingGenes95958_09-mar-2017.xls	10,4 kB	Spreadsheet	mar 9 2017
missingGenes95984_09-mar-2017.xls	15,4 kB	Spreadsheet	mar 9 2017
missingGenes95993_09-mar-2017.xls			

**missingGenes94614\_09-mar-2017.xls:**

A	B	C
1	Homolog Gene	Homolog Product Name
2	2507146062	nitrogenase iron protein NifH
3	2623273787	Mo-nitrogenase iron protein subunit NifH
4	2623430726	Mo-nitrogenase iron protein subunit NifH
5	2582939749	Mo-nitrogenase iron protein subunit NifH (EC 1.18.6.1)
6	2582912046	Mo-nitrogenase iron protein subunit NifH (EC 1.18.6.1)
7	2512782900	nitrogenase iron protein NifH
8	2525223029	Mo-nitrogenase iron protein subunit NifH (EC 1.18.6.1)
9	2518638453	Mo-nitrogenase iron protein subunit NifH
10	643569208	Mo-nitrogenase iron protein subunit NifH (EC 1.18.6.1)
11	2642513352	Mo-nitrogenase iron protein subunit NifH
12	2537645618	nitrogenase iron protein NifH
13	2537176955	nitrogenase iron protein NifH
14	2531218137	nitrogenase iron protein NifH
15	2531212928	nitrogenase iron protein NifH
16	2531208877	nitrogenase iron protein NifH
17	2511472069	nitrogenase iron protein NifH
18	2635236845	Mo-nitrogenase iron protein subunit NifH
19	2587846395	Mo-nitrogenase iron protein subunit NifH (EC 1.18.6.1)
20	2638991531	Mo-nitrogenase iron protein subunit NifH
21	650892485	Mo-nitrogenase iron protein subunit NifH (EC 1.18.6.1)
22	2595582873	nitrogenase iron protein NifH

Bottom status bar: "missingGenes94614\_09-mar-2017.xls" selected (11,2 kB) | Sheet 1 of 1 | 22 rows, 19 columns selected