September 27, 2016

MGM-23 Workshop Presentation: Comparative Genomics for Gene Discovery Using Integrated Microbial Genomes (IMG)

Presenter: Rekha Seshadri, Joint Genome Institute (rseshadri@lbl.gov)

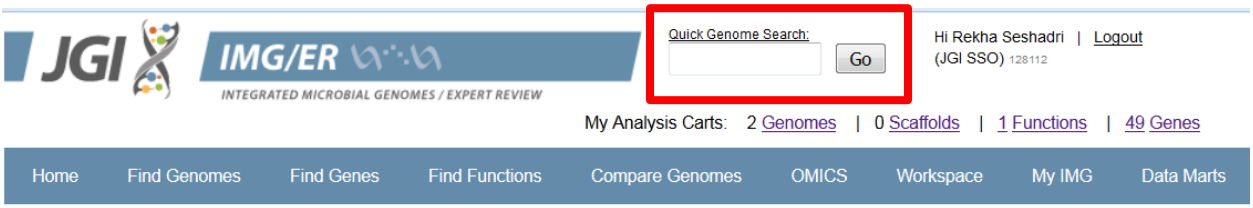
URL: https://img.jgi.doe.gov/cgi-bin/mer/main.cgi

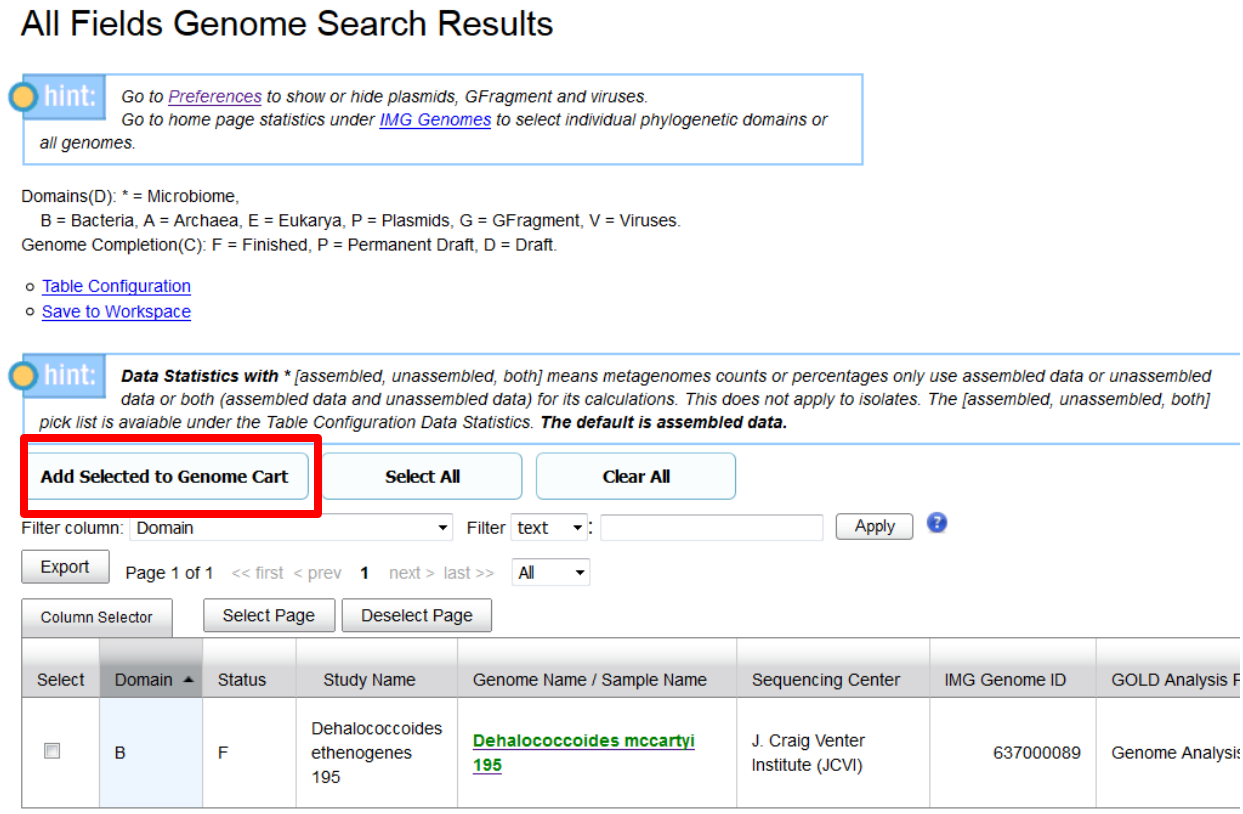
CASE STUDY: Explore and compare two strains of *Dehalococcoides mccartyi* to discover putative gene(s) responsible for complete dechlorination of PCE to nontoxic end product, ethylene.

Strain 195[1]	Strain CBDB1[2]
1.467 Mbp	1.395 Mbp
Chloroethylenes, etc	Chlorobenzenes, etc
PCE->TCE->DCE->VC->Eth	PCE->TCE->DCE
17(+2) RDs	32 RDs

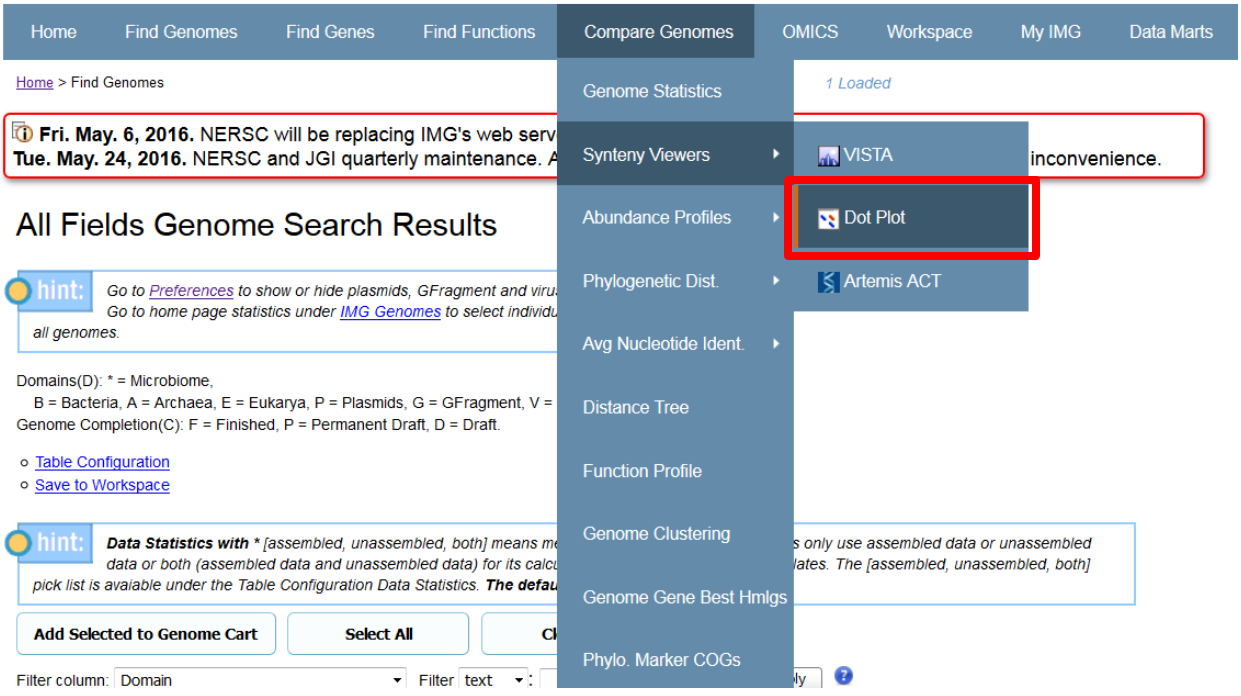
Question 1: Given that these are designated as strains of the same species based on 16S rRNA identity, what is the extent of conservation of gene order or "synteny"?

- *Add the following 2 genomes to your genome cart using the "Quick genome Search" box at the top of the page: Dehalococcoides mccartyi CBDB1 and Dehalococcoides mccartyi 195*

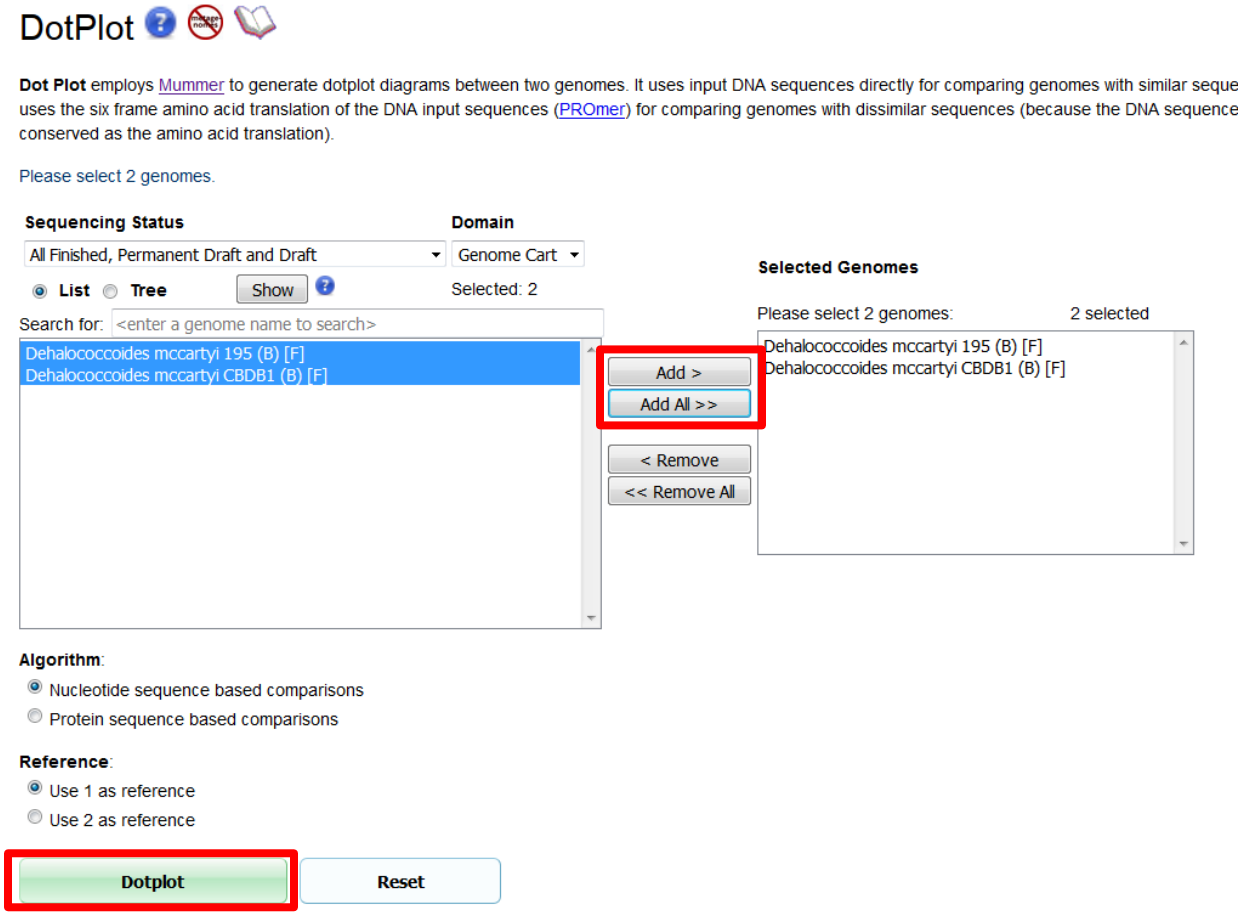




- *Use Compare Genomes > Synteny Viewers > Dot Plot*

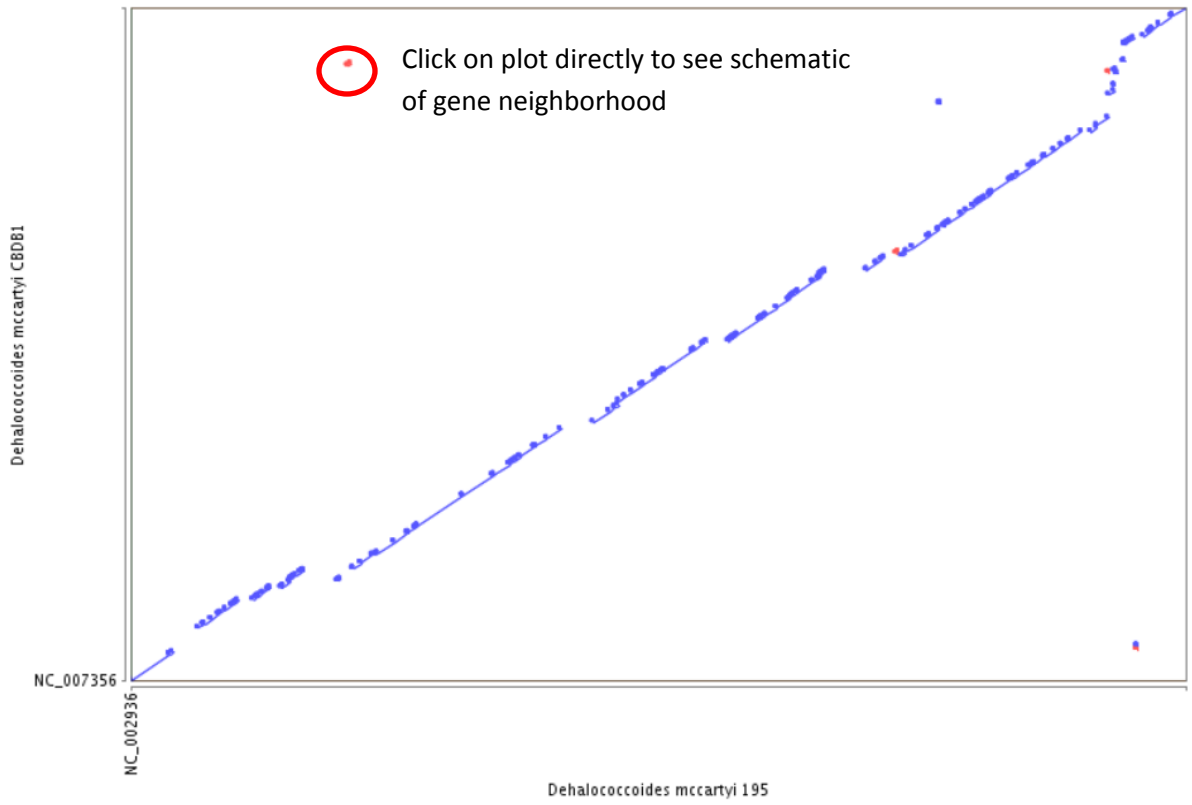


- *Select and Add both genomes and click "Dotplot"*



Dehalococcoides mccartyi 195 vs.
Dehalococcoides mccartyi CBDB1

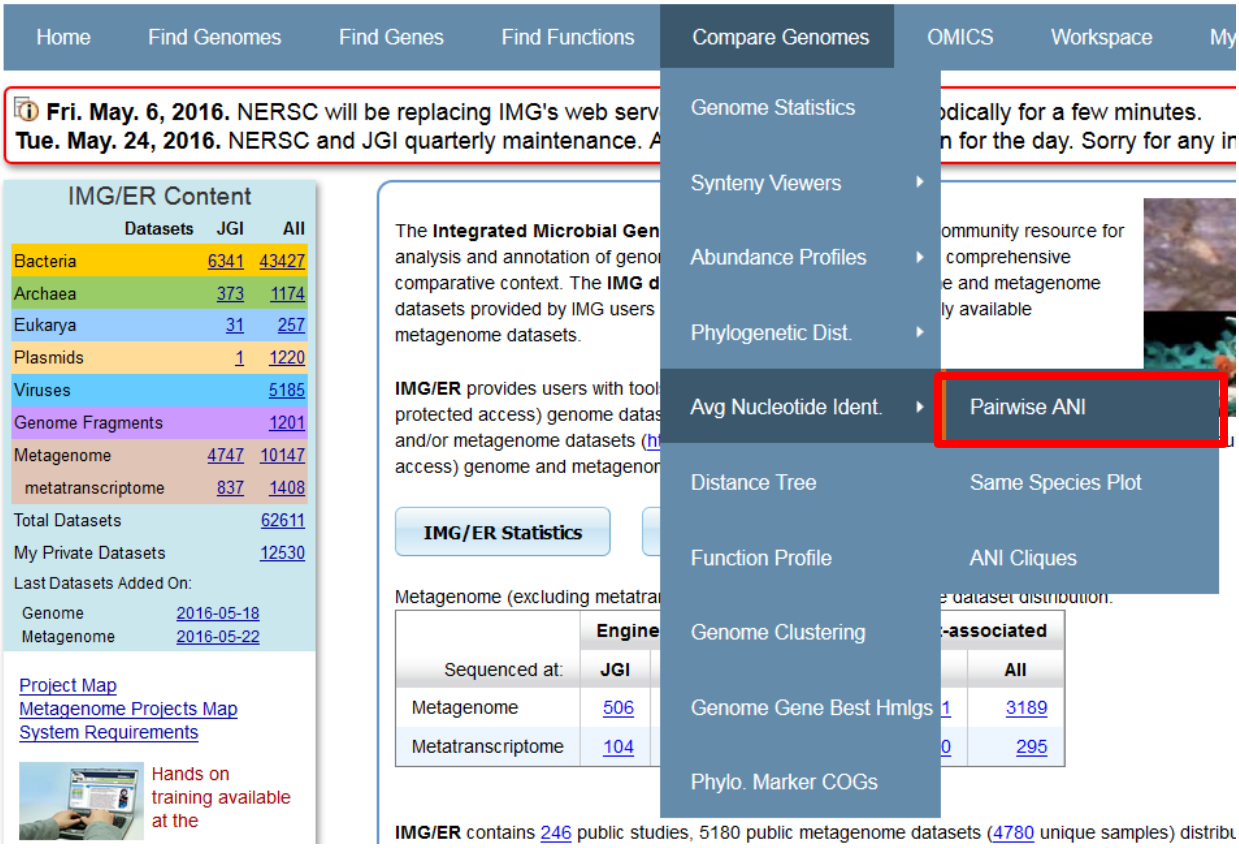
Using nucmer to compare genomes:



Answer 1: Dot plot reveals extensive synteny with some gaps, rearrangement, transposition or inversion appears to involve reductive dehalogenase genes, key enzymes for their hallmark property.

Question 2: What proportion of genes or proteins is conserved?

- *Use Compare Genomes > Average Nucleotide Identity > Pairwise ANI*



Pairwise ANI

BBHs between a genome pair are computed as pairwise bidirectional best nSimScan hits of genes having 70% or more identity and at least 70% coverage of the shorter gene. You may either select genome(s) from IMG or you may upload a nucleotide sequence in FASTA format (using the Upload File button) to compute ANI to selected genome(s) in IMG

Please select up to 100 genomes:

Quick Search: <enter a genome name to search>

Sequencing Status **Domain**

All Finished, Permanent Draft and Draft | Genome Cart

List Tree Show Selected: 1

Search for: <enter a genome name to search>

Dehalococcoides mccartyi 195 (B) [F]
Dehalococcoides mccartyi CBDB1 (B) [F]

Selected Genomes

Pairwise 1: 1 selected

Dehalococcoides mccartyi 195 (B) [F]

Add Upload Sets Remove Upload File

Pairwise 2: 1

Dehalococcoides mccartyi CBDB1 (B) [F]

Add Upload Sets Remove

ANI

Pairwise ANI

Filter column: Genome1 ID Filter text : Apply

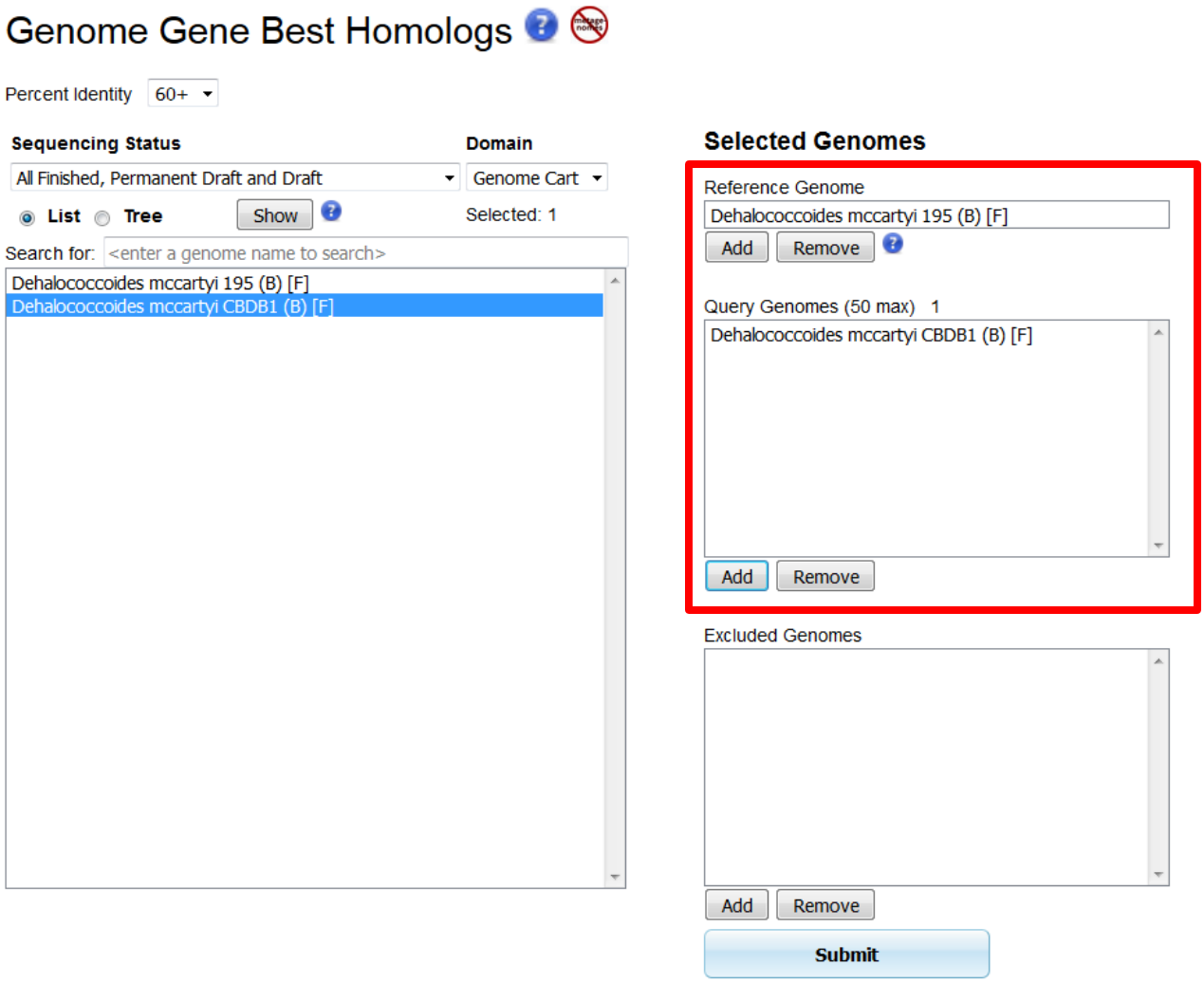
Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector

Genome1 ID	Genome1 Name	Genome2 ID	Genome2 Name	ANI1->2	ANI2->1	AF1->2	AF2->1	Total BBH	Precomputed ?
637000089	Dehalococcoides mccartyi 195	637000090	Dehalococcoides mccartyi CBDB1	86.35	86.34	0.810	0.830	1251	Yes

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

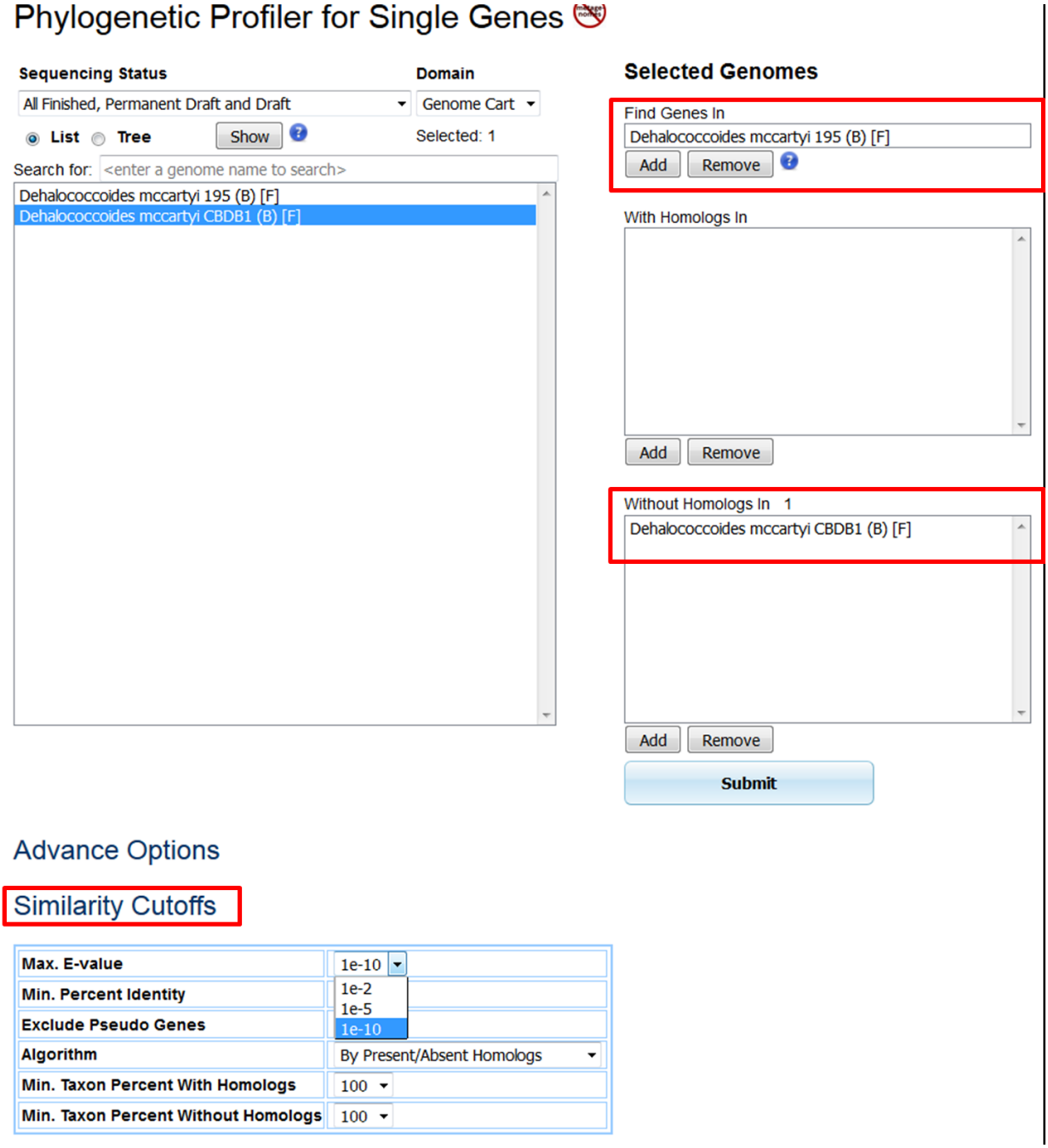
- *To compare proteins, use Compare Genomes > Genome Genes Best Hmlgs*



Answer 2: Both nucleotide and amino acid sequence based comparisons suggest about 75-80% of the genes and proteins are conserved based on best reciprocal hits.

Question 3: How many genes or what functions appear to be unique to Strain 195?

- *Use Find Genes > Phylogenetic Profilers > Single Genes*



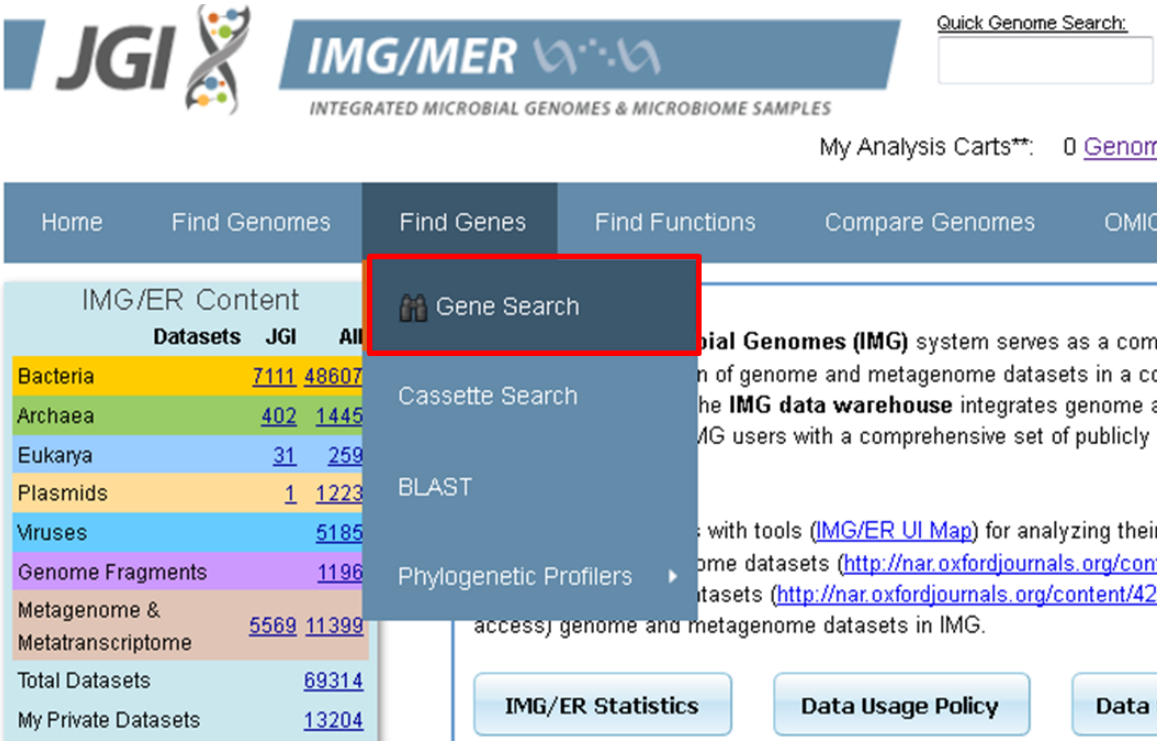
- *Results table filtered by Gene Name for "nitrogenase"*

Select	Result ▲	Gene Object ID	Locus Tag	Gene Name	Length	KEGG Map Name	KEGG Module Name
☐	231	637120711	DET1148	nitrogenase cofactor biosynthesis protein NifB, putative	276	-	-
☐	234	637120714	DET1151	dinitrogenase iron-molybdenum cofactor NifB/Y/X family protein	134	-	-
☐	235	637120715	DET1152	nitrogenase molybdenum-iron protein, beta subunit, putative	451	-	-
☐	236	637120716	DET1153	nitrogenase MoFe cofactor biosynthesis protein NifE	454	-	-
☐	237	637120717	DET1154	Mo-nitrogenase MoFe protein subunit NifK (EC 1.18.6.1)	461	Microbial metabolism in diverse environments Metabolic pathways Chloroalkane and chloroalkene degradation Nitrogen metabolism	Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia
☐	238	637120718	DET1155	nitrogenase molybdenum-iron protein alpha chain	539	Microbial metabolism in diverse environments Chloroalkane and chloroalkene degradation Nitrogen metabolism Metabolic pathways	Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia
☐	239	637120721	DET1158	Mo-nitrogenase iron protein subunit NifH (EC 1.18.6.1)	274	Nitrogen metabolism Metabolic pathways Chloroalkane and chloroalkene degradation Microbial metabolism in diverse environments	Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia Nitrogen fixation, nitrogen => ammonia

Answer 3: About 300 genes appear to be unique to Strain 195 – these include a nitrogenase operon for nitrogen fixation, and a small number of reductive dehalogenases, that might correspond to observed differences in organohalide substrate specificities.

Question 4: The two strains have different dehalogenation profiles (i.e., preferences in halogenated substrate specificities), in particular, *D. mccartyi* strain195 is capable of COMPLETE dechlorination of PCE to ethylene, unlike Strain CBDB1. Can we identify the reductive dehalogenase(s) responsible for the terminal steps?

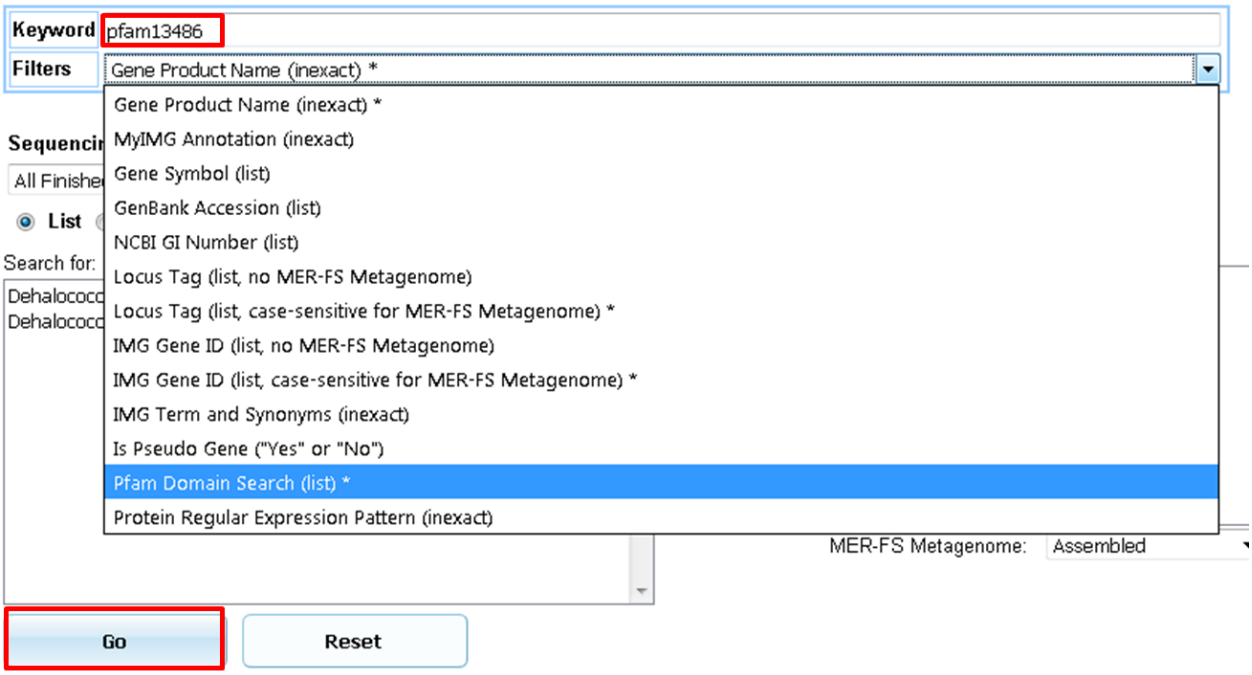
- *Use Find Genes > Gene Search*



- On the next page, select both genomes from your genome cart, use the pfam ID to search for reductive dehalogenases (pfam13486) (why not use the default gene product name search instead? Names can be inconsistent)

Gene Search

Find genes in selected genomes by keyword. It's required to add selections into "**Selected Genomes**" unless blocked.
*MER-FS Metagenome supported search filters.



- Nest, from Gene Cart page, use the Sequence Alignment Tab and "Do alignment"

Gene Cart

Only a maximum of 20000 genes can be in cart.

49 gene(s) in cart

Genes in Cart | Functions | Upload & Export & Save | Chromosome Map | Sequence Alignment | Gene Neighborhoods | Profile & Alignment | Edit

Add Genomes of Selected Genes to Cart | Add Scaffolds of Selected Genes to Cart

Toggle Selected | Select All | Clear All | Remove Selected

49 of 49 rows selected

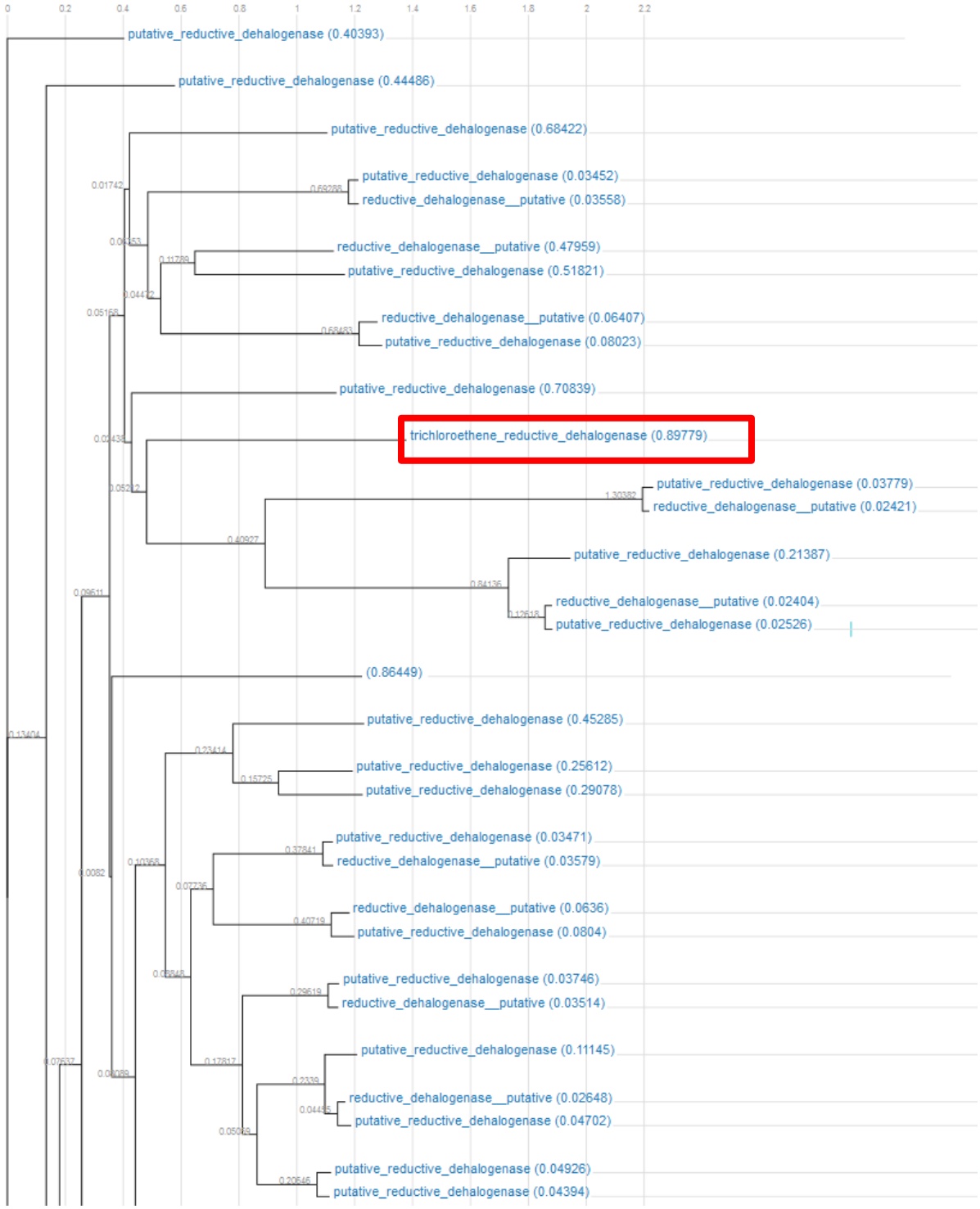
Filter column: Gene Product Name Filter text : Apply

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

Column Selector | Select Page | Deselect Page

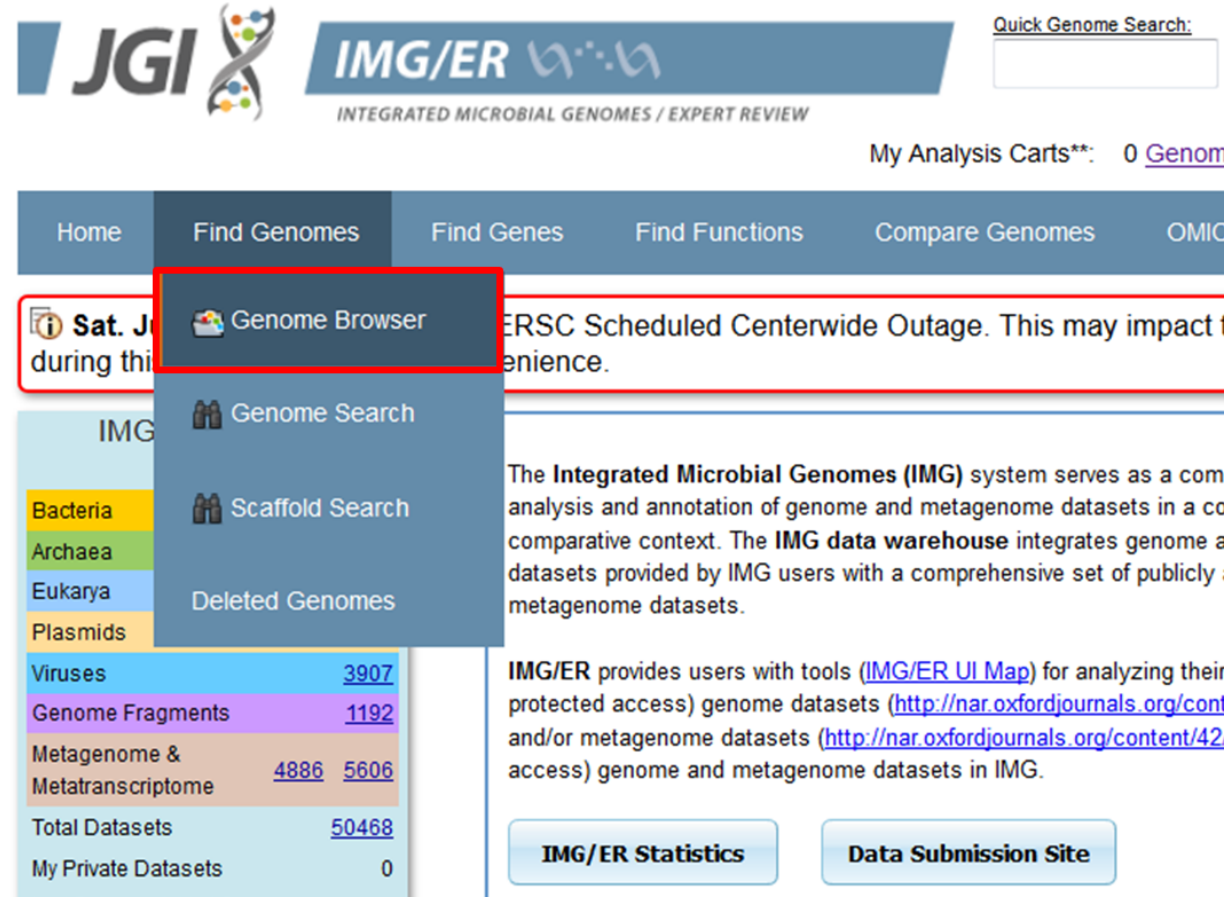
Select	Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name	Batch[1]	Amino Acid Sequence Length (aa)
☑	640733214	DET0162		637000089	Dehalococcoides mccartyi 195	3	486
☑	637703851	cbdb_A1491	putative reductive dehalogenase	637000090	Dehalococcoides mccartyi CBDB1	2	482
☑	637702640	cbdb_A84	putative reductive dehalogenase	637000090	Dehalococcoides mccartyi CBDB1	2	488
☑	637703982	cbdb_A1624	putative reductive dehalogenase	637000090	Dehalococcoides mccartyi CBDB1	2	495
☑	637703944	cbdb_A1582	putative reductive dehalogenase	637000090	Dehalococcoides mccartyi CBDB1	2	491

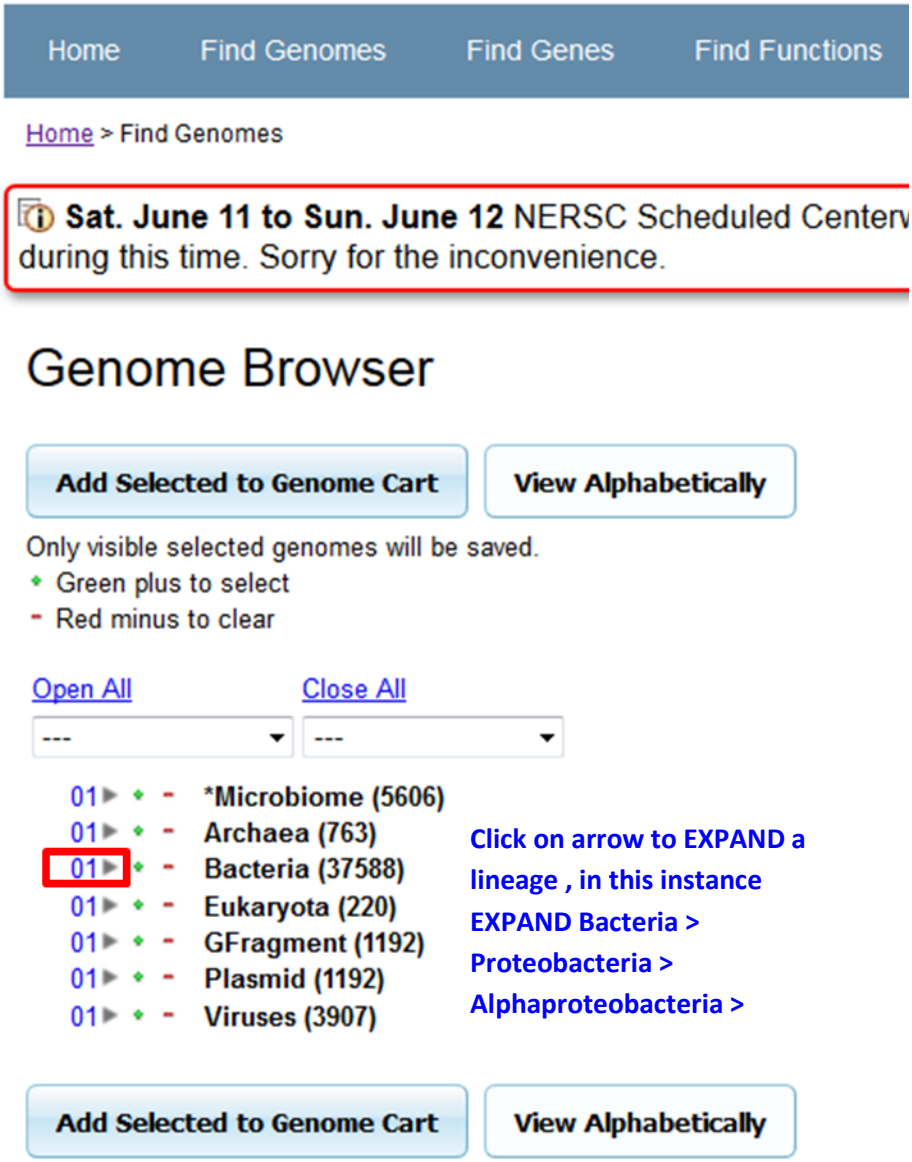
- From the alignment results page, select Rectangular Phylogram" tab to view a tree



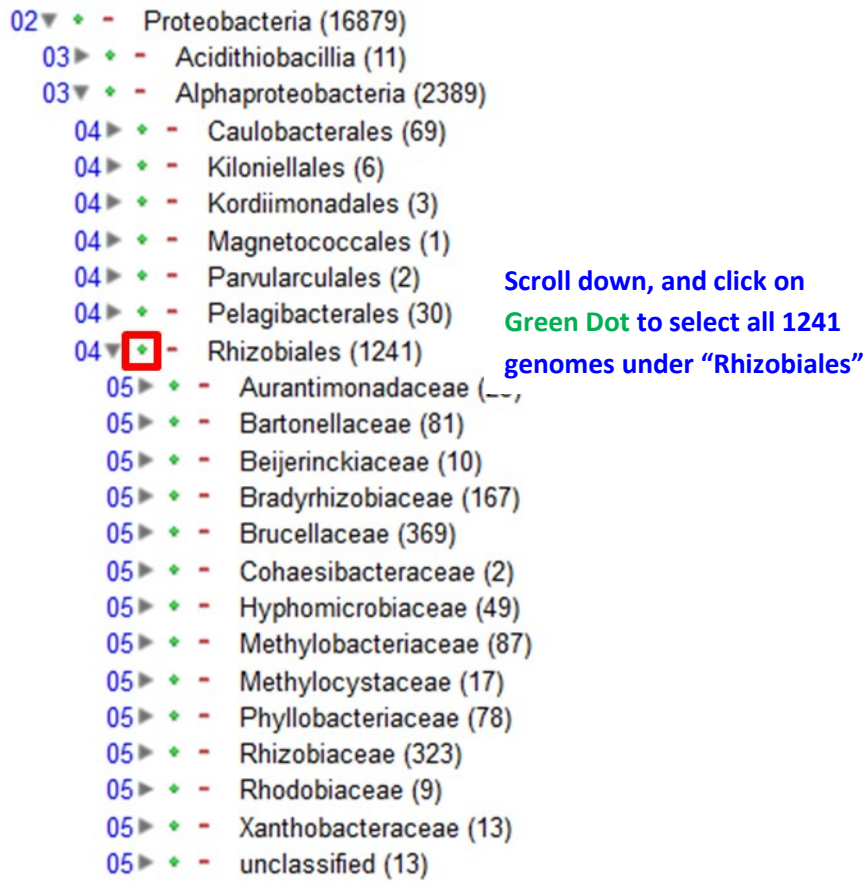
Answer 4: From the tree, it is possible to discern a handful of divergent candidates in strain 195 relative to CBDB1– the gene described as "trichloroethene reductive dehalogenase" is experimentally validated to dehalogenate DCE and VC to ethylene[4]

CASE STUDY: Discovering novel determinants of symbiosis through comparative genomics of 100+ root nodule bacteria against a control genomes set of non-plant associated bacteria. The objective is to showcase the metadata available in IMG courtesy of the GOLD effort[5].



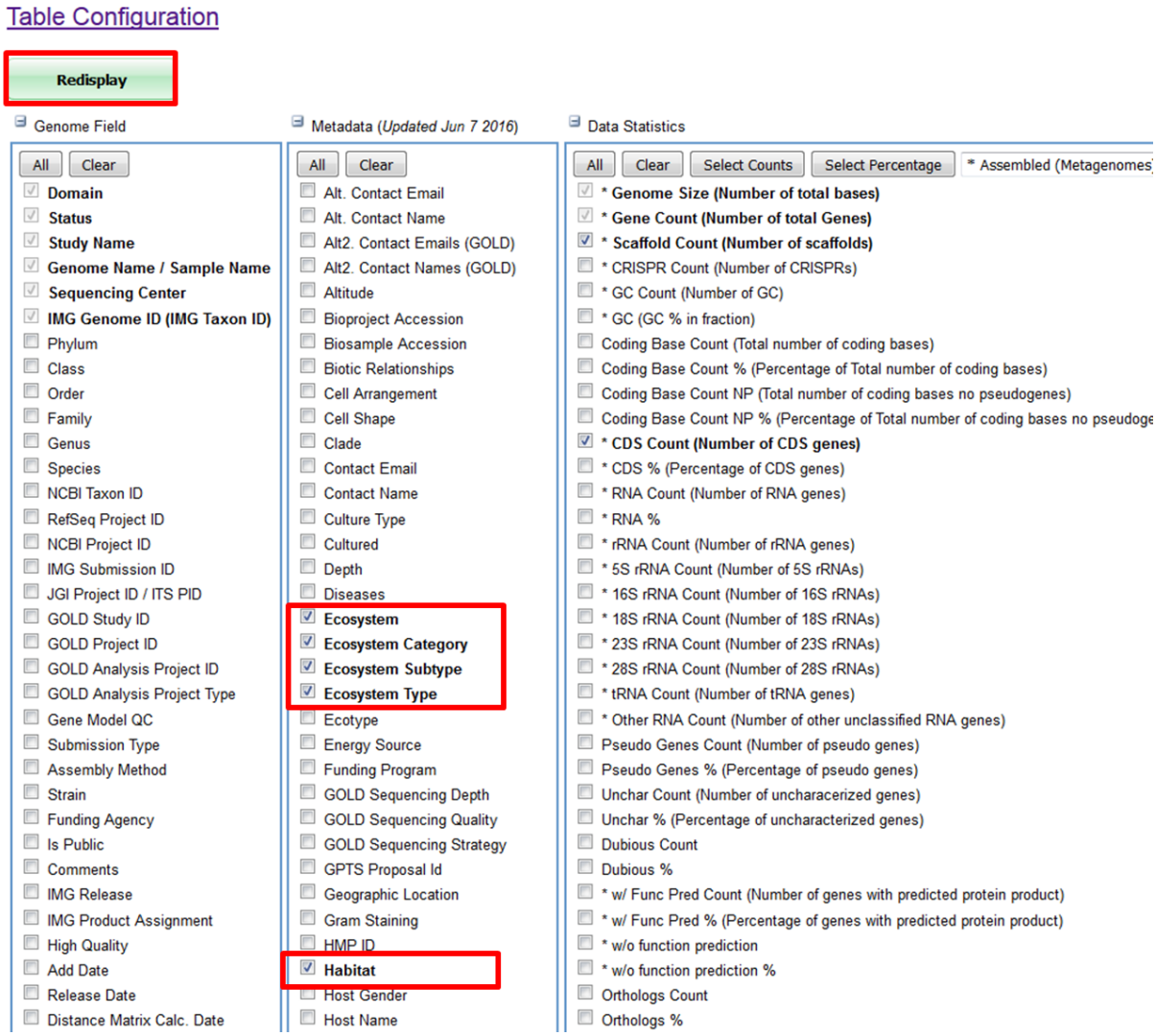

Home
Find Genomes
Find Genes
Find Functions
Home > Find Genomes
Sat. June 11 to Sun. June 12 NERSC Scheduled Centerv
during this time. Sorry for the inconvenience.
Genome Browser
Add Selected to Genome Cart
View Alphabetically
Only visible selected genomes will be saved.
Green plus to select
Red minus to clear
Open All
Close All

---
01 *Microbiome (5606)
01 Archaea (763)
01 Bacteria (37588)
01 Eukaryota (220)
01 GFragment (1192)
01 Plasmid (1192)
01 Viruses (3907)
Click on arrow to EXPAND a lineage , in this instance EXPAND Bacteria > Proteobacteria > Alphaproteobacteria >
Add Selected to Genome Cart
View Alphabetically



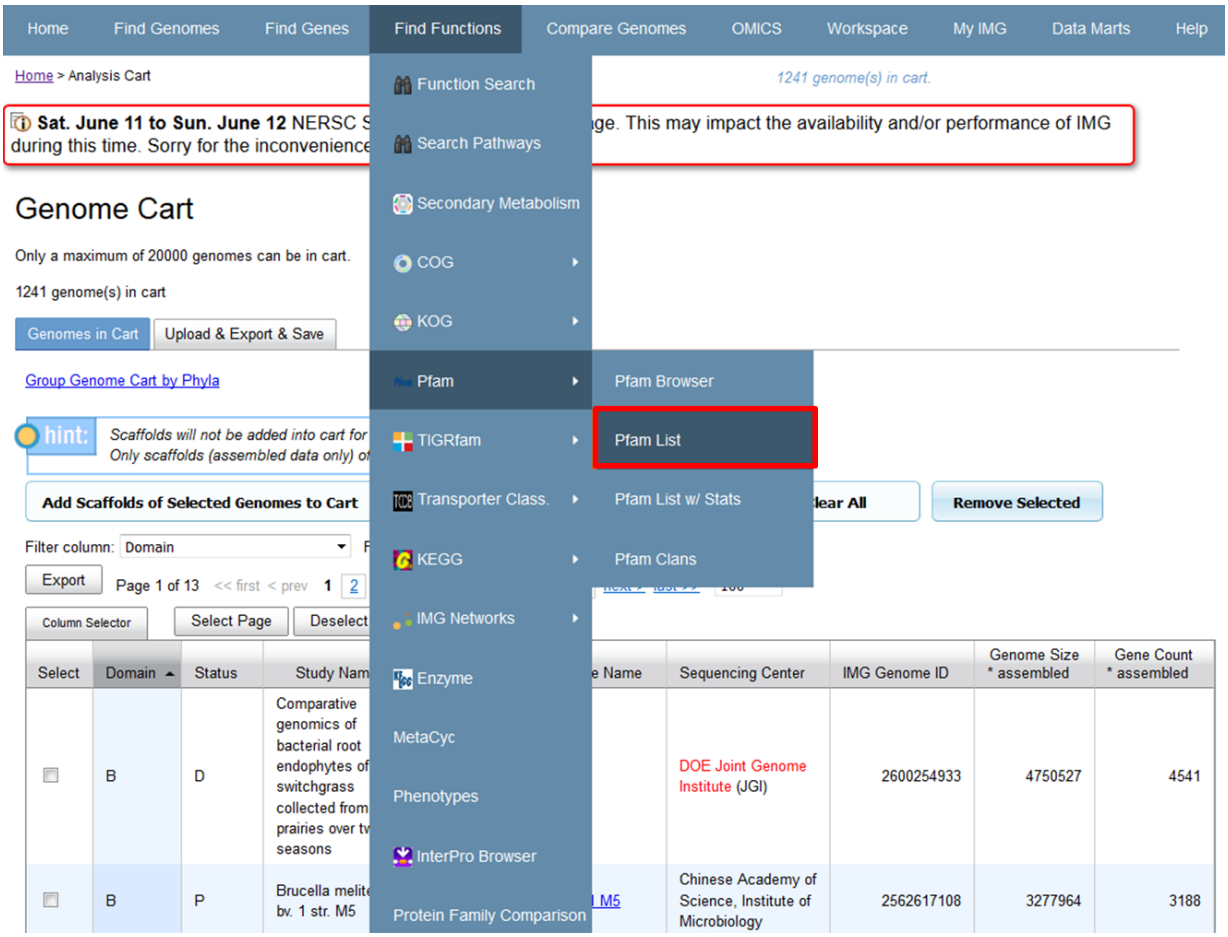
- When the next screen loads, add these 1241 selected genomes to Genome Cart:

- 1241 genomes are now loaded in your genome cart. On this page, scroll down to "Table configuration" and make your "metadata "selections and Redisplay the table:

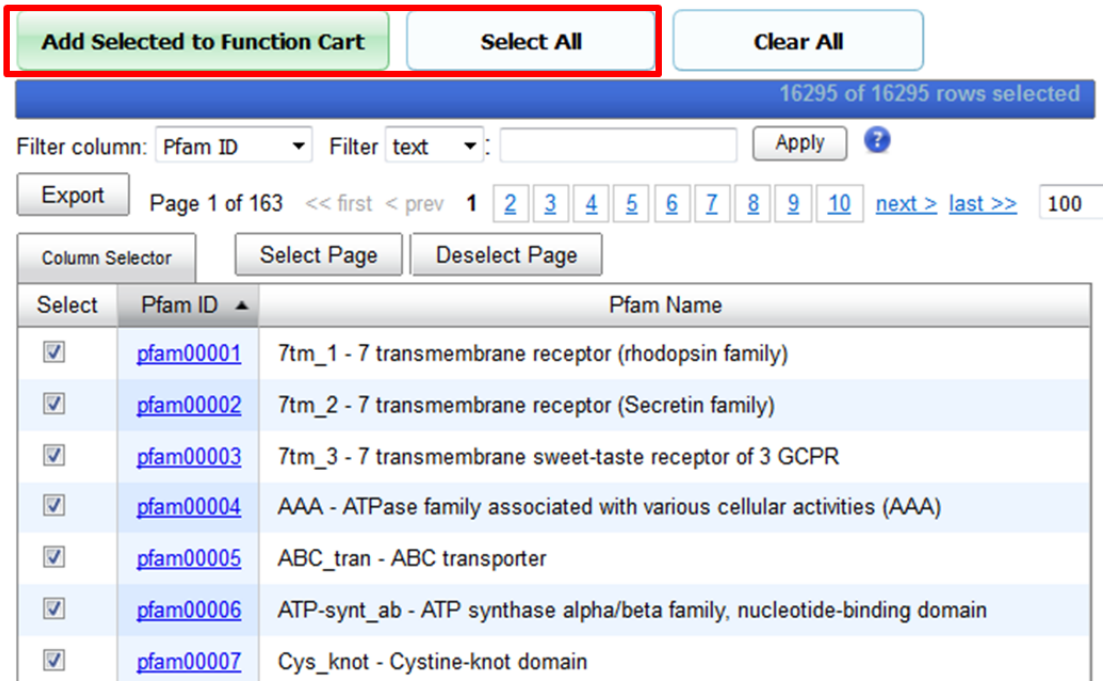


- Curate your list of “Control” genomes (genomes from within the Order Rhizobiales with NO PLANT ASSOCIATION) using these “metadata “

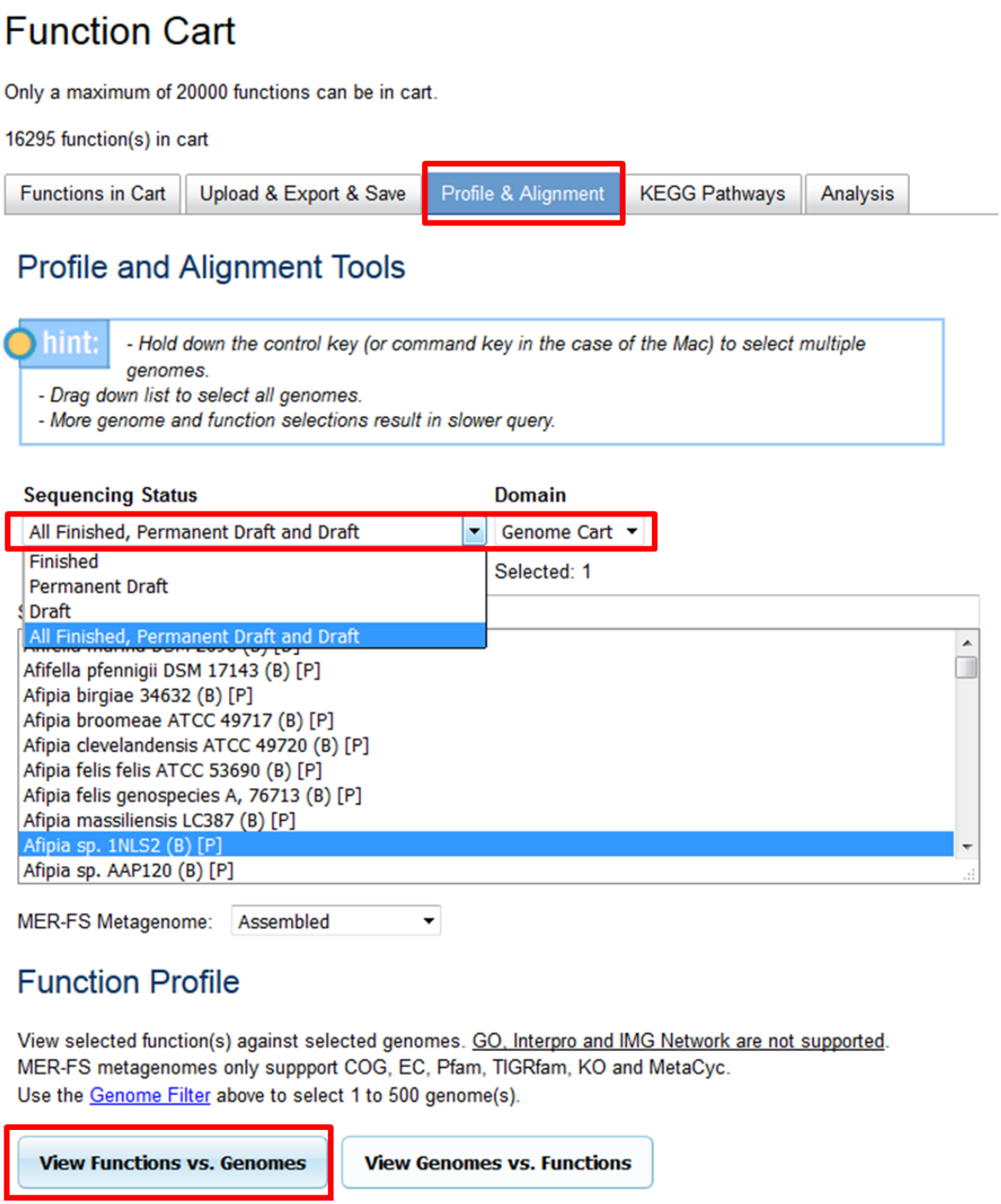
- Next, retrieve gene counts of all Pfams for all Control genomes. First use Find Functions > Pfam > Pfam list > Select All 16,295 pfams and Add to Function Cart



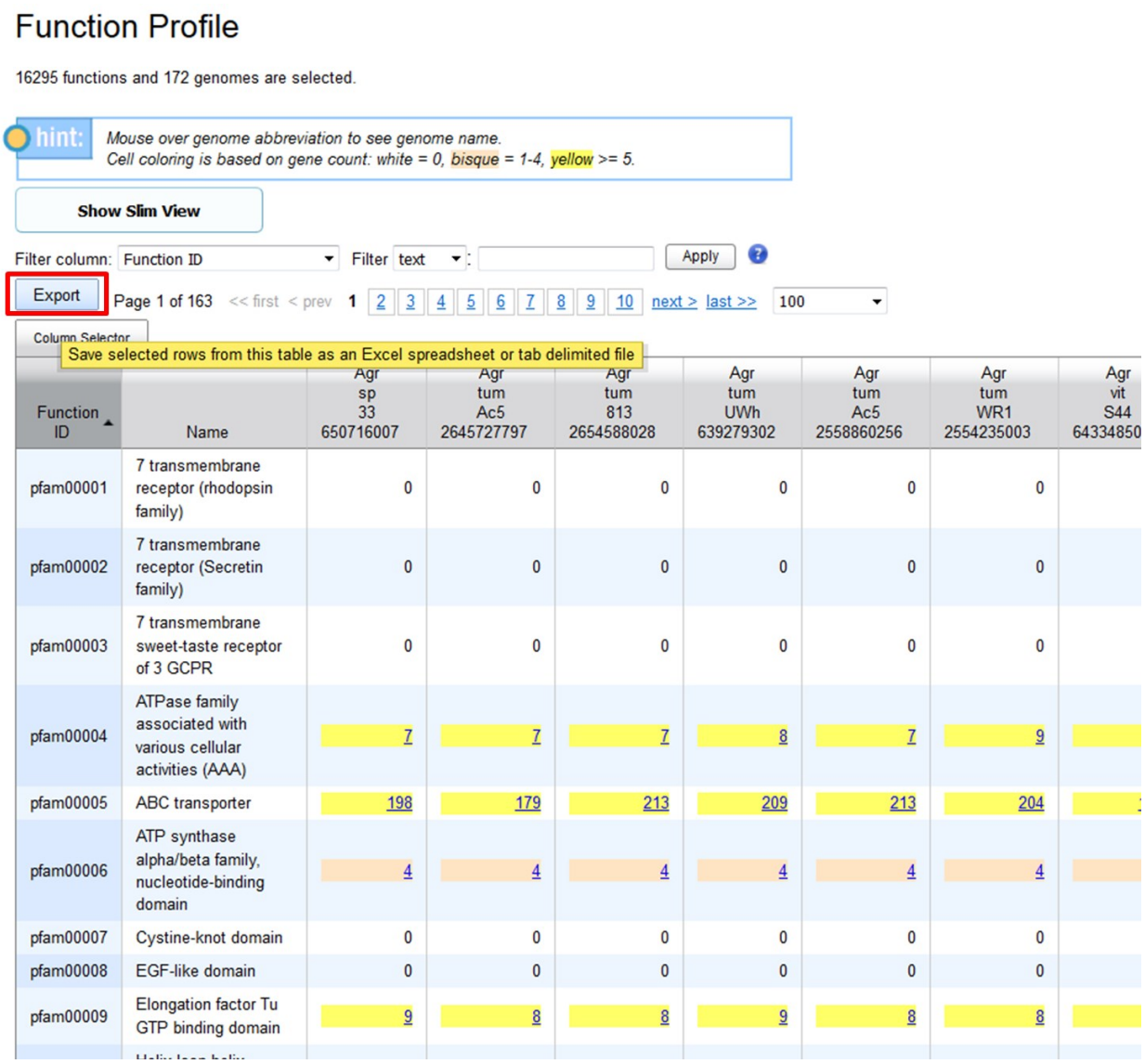
Pfam Families



- Tab over to “Profile & Alignment” on the Function Cart page, populate with genomes from “Genome Cart” and retrieve gene counts for all genomes using “View Genomes vs. Functions”



- Export the Function Profile table for further analyses.



Function Profile

16295 functions and 172 genomes are selected.

hint: Mouse over genome abbreviation to see genome name.
Cell coloring is based on gene count: white = 0, bisque = 1-4, yellow >= 5.

Show Slim View

Filter column: Function ID Filter text : Apply

Export Page 1 of 163 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector

Save selected rows from this table as an Excel spreadsheet or tab delimited file

Function ID	Name	Agr sp 33 650716007	Agr tum Ac5 2645727797	Agr tum 813 2654588028	Agr tum UWh 639279302	Agr tum Ac5 2558860256	Agr tum WR1 2554235003	Agr vit S44 64334850
pfam00001	7 transmembrane receptor (rhodopsin family)	0	0	0	0	0	0	
pfam00002	7 transmembrane receptor (Secretin family)	0	0	0	0	0	0	
pfam00003	7 transmembrane sweet-taste receptor of 3 GCPR	0	0	0	0	0	0	
pfam00004	ATPase family associated with various cellular activities (AAA)	7	7	7	8	7	9	
pfam00005	ABC transporter	198	179	213	209	213	204	
pfam00006	ATP synthase alpha/beta family, nucleotide-binding domain	4	4	4	4	4	4	
pfam00007	Cystine-knot domain	0	0	0	0	0	0	
pfam00008	EGF-like domain	0	0	0	0	0	0	
pfam00009	Elongation factor Tu GTP binding domain	9	8	8	9	8	8	

Results: Results from this exercise, overall strategy, and more, are described in the following publication[6]: http://www.nature.com/articles/srep16825

REFERENCES

1. Genome sequence of the PCE-dechlorinating bacterium Dehalococcoides ethenogenes. Seshadri R, et al. Science, 2005 Jan 7. PMID 15637277

2. Genome sequence of the chlorinated compound-respiring bacterium Dehalococcoides species strain CBDB1. Kube M et. Al. Nat Biotechnol. 2005 Oct;23(10):1269-73.

3. Microbial species delineation using whole genome sequences, Varghese, N. et. al. Nucl. Acids Res. (2015)

4. Trichloroethene reductive dehalogenase from Dehalococcoides ethenogenes: sequence of tceA and substrate range characterization. Magnuson JK, et al. Appl Environ Microbiol, 2000 Dec. PMID 11097881

5. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Reddy TBK, et. al., Nucl. Acids Res. (2014)

6. Discovery of Novel Plant Interaction Determinants from the Genomes of 163 Root Nodule Bacteria. Seshadri, R. et.al. Nat. Sci. Rep. 2015. 5:16825.