Welcome to SEED Tutorial.

Assignment overview

Part I. Find a gene. Locate specific gene/protein page in SEED dedicated to the protein assigned to you. Familiarize yourself with different types of the data, tools, and links to external resources available on a gene/protein page

Part II. Annotate a gene. Apply SEED tools (utilizing straightforward homology-based projections, as well as functional and genome context analysis) to reject, refine, or confirm current functional annotation of "your" protein.

Part III. Find a subsystem. Identify functional Subsystem(s) in SEED that your protein potentially belongs to.

Part IV. Explore a subsystem. Learn the main tools of subsystem visualization and analysis.

Part V. Expand a subsystem. Learn the tools for subsystem projection onto additional genomes. Project this Subsystem onto the microbial genome of your protein. Predict how this pathway (subsystem) is implemented in this particular organism.

Detailed Assignment

PART 0. Enter SEED using one of the three URLs:

http://theseed.uchicago.edu/FIG/index.cgi	if your terminal is numbered 1 through 7
http://neisseria.uchicago.edu/FIG/index.cgi	if your terminal is numbered 8 through 16
http://shigella.nmpdr.org/FIG/index.cgi	if your terminal is numbered 17 through 25

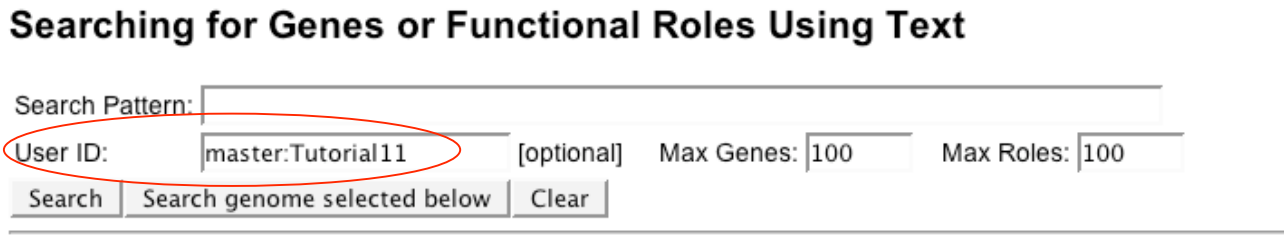
PART I. FIND A GENE.

Every protein-encoding gene (PEG) from every genome in SEED has an individual WEB page containing a variety of data about the protein and the corresponding gene, as well as a number of tools for protein annotation and analysis.

I.1. To identify a protein assigned to you open the following WEB page in a new Tab of your WEB Browser:
http://www-unix.mcs.anl.gov/SEEDWiki/admin/moin.cgi/Lightweight_20SEED_20Tutorial
This is a SEED WIKI page with the complete list of assignments.

I.2. Out of 20 proteins on this list pick the one, which number matches the number on your computer terminal

I.3. In order to annotate genes (Part 2 and beyond), you need to authenticate yourself in the box "User ID" under the caption:



Please type your user ID exactly as it appears on the Assignment page under your Terminal number (master:Tutorial##). Make sure that you use the same username throughout the tutorial.

I.4. Copy protein sequence or ID (your choice) and paste into an appropriate window on the main **FIG search** page (opened in the first Tab of your Browser):

(i) If you chose to copy an ID, paste it (just the number, omit "gi|"), into the window **Searching for Genes or Functional Roles Using Text**, and press **Search** button. To limit your search to the genome of your protein - scroll down the page, highlight this genome, scroll back up and click **Search genome selected below** button.

(ii) If you chose to copy a sequence, paste it into the window **Searching DNA or Protein Sequences (in a selected organism),** scroll up the page, highlight the genome of your protein, scroll back down and click **Search for matches** button (check that **Search Program** is set for "blastp")

Both searches should generate a single PEG ID or a list of IDs matching your search criteria. A complete PEG ID in SEED looks something like that: **fig|562.2.peg.1246**,
where "fig|562.2" is a genome ID and version (number after the dot) and "peg.1246" is an ID of a specific protein in this genome

I.5. Click on the PEG ID to follow the link to the corresponding PEG page

PART II. ANNOTATE A GENE.

II.1. Browse the PEG page to learn basic characteristics of your protein, the immediate genomic neighborhood of the corresponding gene, its current annotation in SEED, as well as in other major genomic databases, check out links to external resources.

II.2. Each PEG page begins with a table we call the context. It represents the region on the chromosome (or fragment of a chromosome that we often call a contig). The first column in this table has the label fid, which stands for feature ID. The start and end columns give the exact coordinates of the gene on the contig (not including the stop codon). The size is in bases. The strand is + or -, and the gap is the distance between two genes (genes that overlap have negative values for the gap, which is something worth checking occasionally). The next two columns, **fc** and **neigh**, are important. The genes with a FC in the **fc** column appear to have some evidence supporting the hypothesis that they tend to co-occur with the gene you are positioned on (see notes on **F**unctional **C**oupling below). The **neigh** column will be marked for genes that are known to play closely-related functional roles (e.g., occurring in the same pathway), this evidence is not kept up to date, however.

II.3. Examine graphical depiction of the chromosomal region. The meanings of the colors are as follows:

-- you are positioned on the green gene,
-- the red genes are apparently unrelated genes, and
-- blue genes are genes that might be functionally related (there is some evidence based on co-occurrence close to the given gene in several genomes).

II.4. To generate a list of similarities, containing instances of this gene and its close homologs in other genomic databases as well as in SEED - scroll down to the bottom of the PEG page:

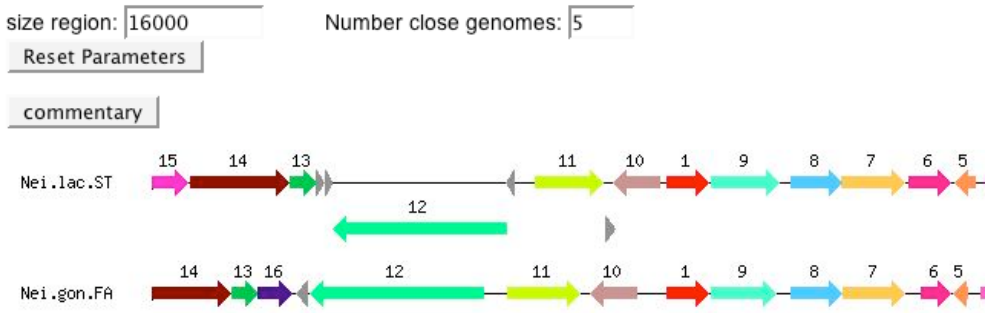
Click **Similarities** button. Analyze the resultant table of homologs. You will likely see multiple instances of the same proteins with annotations coming from several public archives (note protein IDs characteristic of GenBank. UniProt, KEGG)

II.5. To generate a non-redundant list of similarities containing only homologs of the query protein amenable for annotation in the SEED database (all protein IDs will have a form "fig___._peg___") choose "**Just FIG IDs (all)"** from the drop-down menu, type 50 in "max expand" box, and click **Similarities** button:

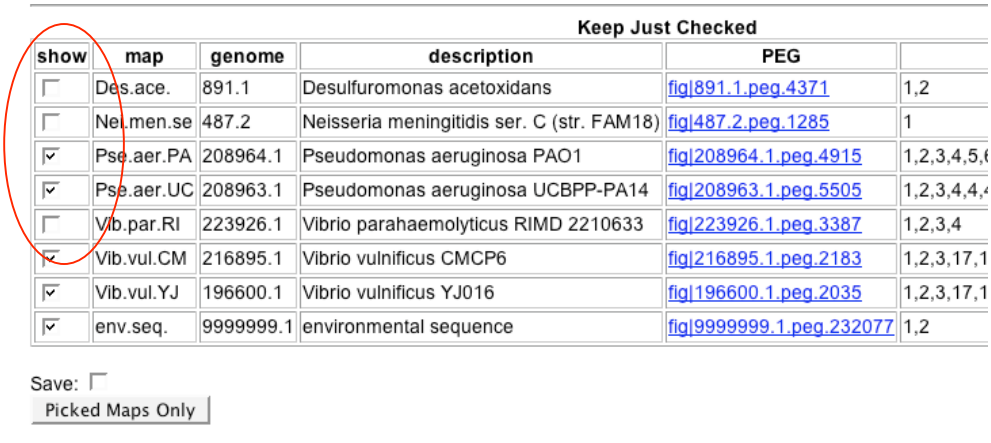
Analyze the resultant table of homologs.
Feel free to consult **Help with SEED similarities options** available on each PEG page for a detailed explanation of this tool.

II.6. Tool **To compare region** allows you to explore chromosomal neighborhoods of the closest homologs of your gene in other genomes.

The output will show -- in addition to the genes in the genome you are examining – the corresponding regions in closely related genomes "pinned" around homologs of your gene. The query gene is in red, numbered as 1. Other homologous genes in the region are shown by arrows with matching colors and numbers. Genes not conserved within the region are colored gray. Mouse over each arrow for more details. You can expand the area shown or change the number of genomes included in the display by typing new parameters into the corresponding boxes and pressing Reset Parameters button:



Pressing Commentary button will activate a page listing groups of homologs in the order of their abundance (#1 = present in all the genomes included in the display, #2 – present in most of the genomes, ... #15 – present in just a few genomes). This page can be used for annotation. Its unique advantage over straight-forward homology projections is in considering chromosomal context: only close homologs in similar chromosomal clusters are assigned with identical specific functions. The table at the very bottom of the Commentary page explains genome abbreviations and lets you pick specific genomes to be included in display. Use check-boxes in **show** column on the left to select genomes of interest, then press Picked Maps Only button at the bottom:



Keep Just Checked

show	map	genome	description	PEG	
☐	Des.ace.	891.1	Desulfuromonas acetoxidans	fig|891.1.peg.4371	1,2
☐	Nei.men.se	487.2	Neisseria meningitidis ser. C (str. FAM18)	fig|487.2.peg.1285	1
☑	Pse.aer.PA	208964.1	Pseudomonas aeruginosa PAO1	fig|208964.1.peg.4915	1,2,3,4,5,6
☑	Pse.aer.UC	208963.1	Pseudomonas aeruginosa UCBPP-PA14	fig|208963.1.peg.5505	1,2,3,4,4,4
☐	Vib.par.RI	223926.1	Vibrio parahaemolyticus RIMD 2210633	fig|223926.1.peg.3387	1,2,3,4
☑	Vib.vul.CM	216895.1	Vibrio vulnificus CMCP6	fig|216895.1.peg.2183	1,2,3,17,1
☑	Vib.vul.YJ	196600.1	Vibrio vulnificus YJ016	fig|196600.1.peg.2035	1,2,3,17,1
☑	env.seq.	9999999.1	environmental sequence	fig|9999999.1.peg.232077	1,2

Save: ☐
Picked Maps Only

A modified "compare region" display will be generated. If necessary, it can be fine-tuned even further by changing similarity threshold (optional):

II.7. Functional Coupling

Functionally related genes tend to cluster on prokaryotic chromosomes. This fact is the basis for a number of techniques used to gain clues relating to the function of hypothetical proteins. The genes with **FC** in the **fc** column in the context table appear to have some evidence supporting the hypothesis that they tend to co-occur with the gene you are positioned on. Click on the **FC** link. This produces a visual depiction of the co-occurrences. Although it looks very similar to the output of the **Compare region** tool, please keep in mind that the two illustrate very different things: while Compare region simply shows you immediate neighborhoods of the closest homologs of your gene in other genomes, regardless of the presence/absence of any other genes in its vicinity; the FC display includes only those genomes in which 2 "functionally coupled" genes are co-localized or "clustered" on the chromosome.

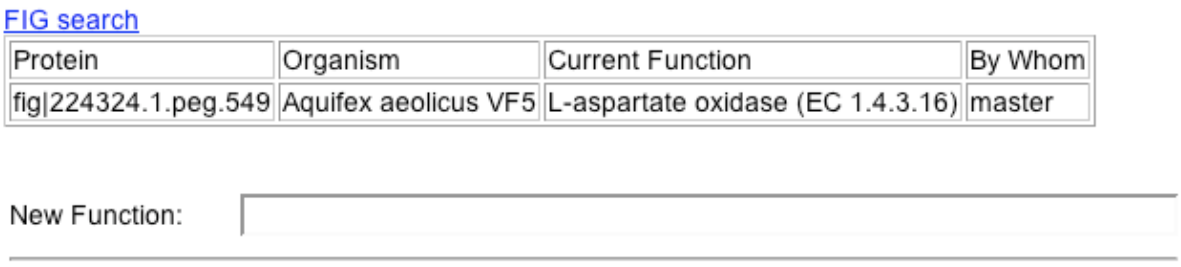
II.8. The detailed analysis of your protein described above has probably provided some keys to its potential function. You can annotate your PEG now to reflect this knowledge. Keep in mind that in order for a PEG in SEED to be connected to a Subsystem it's annotation must exactly match the name of the corresponding functional role as it appears in the table of Functional Roles in a SS.

There are multiple ways to annotate PEGs in SEED. Choose one of the 3 strategies described below to annotate your protein:

(i) Annotating from a PEG page:

-- click "To Make an Annotation" link

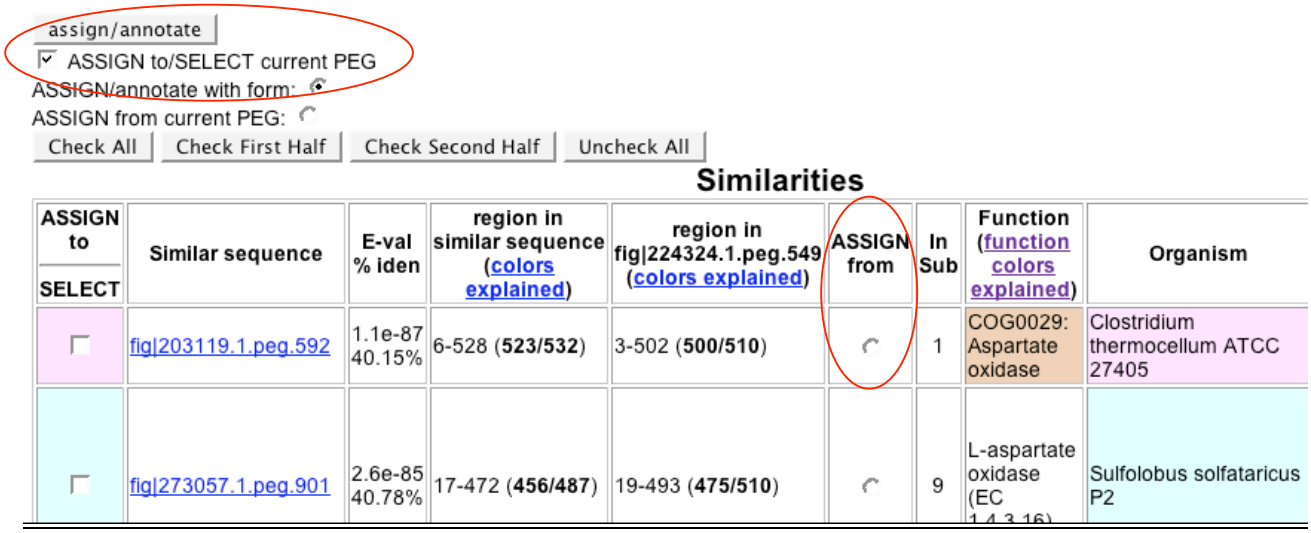
-- a form will appear:



-- type new annotation into "New Function" window and press add annotation button (note that to avoid typos, it is better to copy the corresponding functional role from your SS and paste it into this window)

-- a report is generated in a new window – it is safe to close it

(ii) Annotating from Similarities table:



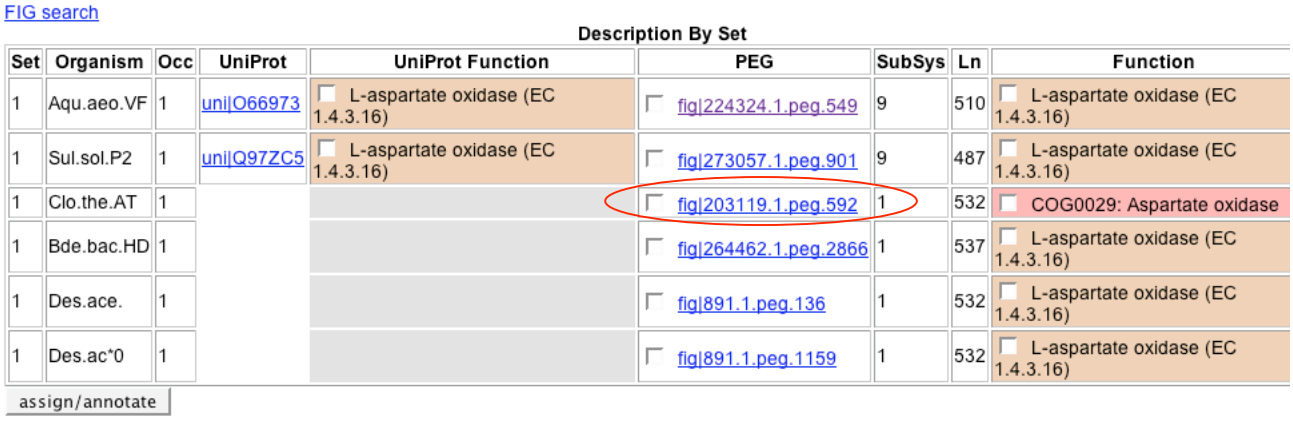
-- click on a round check-box in the **ASSIGN from** column near a homolog with specific annotation you want to propagate

-- make sure "**ASSIGN to/SELECT current PEG**" check-box right above the SIMs table is checked

-- press assign/annotate button

-- the current PEG is re-annotated, a report page is generated in a new window - it is safe to close it

(iii) Annotating from a Commentary page:



-- use **To Compare Region** tool to generate Commentary page as described in II.6

-- in the PEG column check the PEG you intend to re-annotate

--chose "Function" to annotate it from: you can use either the right-most column (**Function**) or **UniProt Function** column. Check the box near the function you want to propagate

-- press assign/annotate button

-- your PEG will be re-annotated to match the source you selected

-- close report page

Part III. FIND A SUBSYSTEM.

There are several ways you may use to find a subsystem(s) that involves a functional role (assignment or annotation) of "your" protein :
(i) If your protein has been already included in one (or more) of subsystems, you should see a link on it's PEG page pointing to this subsystem:

Subsystems in which this peg is present

Subsystem	Role
NAD_and_NADP_tutorial	Quinolinate synthetase (EC 4.1.99.-)

(ii) If this is not the case, check if any homologs of your protein may have been included in a subsystem.' Such protein(s) in Similarity tables will have a numerical entry (1, 2, etc) in the column "In Sub", indicating the number of different Subsystems this PEG is a part of:

Similarities

ASSIGN to SELECT	Similar sequence	E-val % iden	region in similar sequence (colors explained)	region in fig|187410.1.peg.3003 (colors explained)	ASSIGN from	In Sub	Function (function colors explained)	Organism
☐	fig|229193.1.peg.998	0 100.00%	1-353 (**353/353**)	1-353 (**353/353**)	○		Quinolinate synthetase (EC 4.1.99.-)	Yersinia pestis biovar Medievalis str. 91001
☐	fig|214092.1.peg.1256	0 100.00%	1-353 (**353/353**)	1-353 (**353/353**)	○	1	Quinolinate synthetase (EC 4.1.99.-)	Yersinia pestis CO92

Go to the respective PEG page (by clicking on its ID) and then follow the subsystem link as described above.

(iii) You may use the section on the SEED Entry Page: "Locate PEGs in Subsystems" to search for a relevant subsystem using EC#, function name (if you are lucky) and protein ID (follow instructions on the Entry Page)

(iv) Finally, you may browse a list of subsystems in SEED (or use your browser's "find in page" functionality) for a potentially relevant term (eg NAD biosynthesis, etc). Reach the list of subsystems by clicking on "Work on Subsystem" button within a Section "Work on Subsystems Using New, Experimental Code" of the SEED Entry page.

Part IV. EXPLORE A SUBSYSTEM:

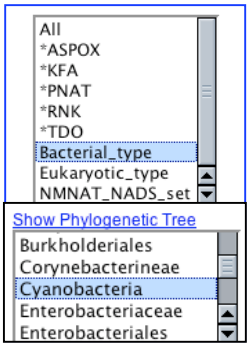
IV.1. Browse the Subsystem (SS) page. It opens with the Table of **Functional Roles** constituting this SS. The roles are defined by the most standard descriptive names, for example enzyme names and corresponding Enzyme Classification (EC) numbers, whenever they are available. These names must exactly match the gene annotations in the underlying database. **Abbreviations** of functional roles are used in Subsystem Spreadsheet below and in its graphic map.

IV.2. Subsets of Roles table. The concept of sub-sets plays an important role in subsystems encoding and interpretation. They usually represent the most compact units, such as multi-subunit complexes, or variants of pathways. Examples of sub-sets are: "Bacterial type" or "Eukaryotic type". A star (*) in front of a sub-set abbreviated name causes all the functional roles grouped in it

to collapse into a single column in a Subsystem spreadsheet – a useful feature for displaying synonymic functional roles or subunits of multi-subunit complexes

IV.3. Subsystem spreadsheet is simply a table, in which each column represents a functional role in the subsystem, each row represents a specific genome, and cell are populated with proteins that implement functional roles in each organism. Protein IDs are linked to the corresponding PEG pages in SEED

A small set of tools located immediately under the **Subsets of Roles** table allow the reduction of spreadsheet display to a selected sub-set of functional roles and/or to a selected group of organisms. Try using them →



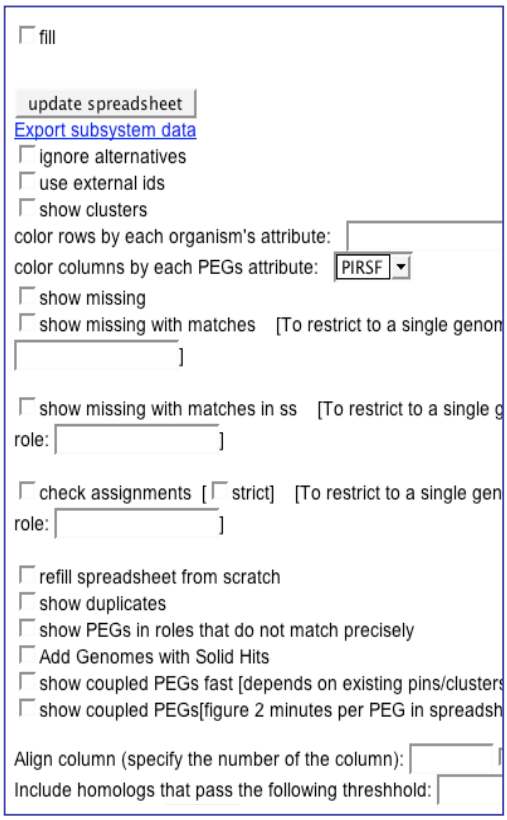
The main Subsystem visualization/construction tools are located on SS page below the SS Spreadsheet. Try using the following:

(i) sorting. Select option "by_phylo" and press update spreadsheet button below. The organisms in the Spreadsheet will be rearranged according to their phylogeny. Selecting "by_pattern" arranges organisms according to the presence/absence of PEGs in the cells of a spreadsheet – a useful tool in analyzing variations in SS implementation in different organisms

(ii) show clusters. Check the box near "show clusters" and press update spreadsheet button. The cells containing proteins clustered on a chromosome will be highlighted by a matching color.

(iii) color rows by each organism's attribute. Choose an attribute from the drop-down menu (e.g. "motile") and press update spreadsheet. See what happens. Legend explaining color usage will also appear.

(iv) color columns by each PEG's attribute – only one protein attribute, namely membership in PIR protein families, can currently be graphically displayed using this option.

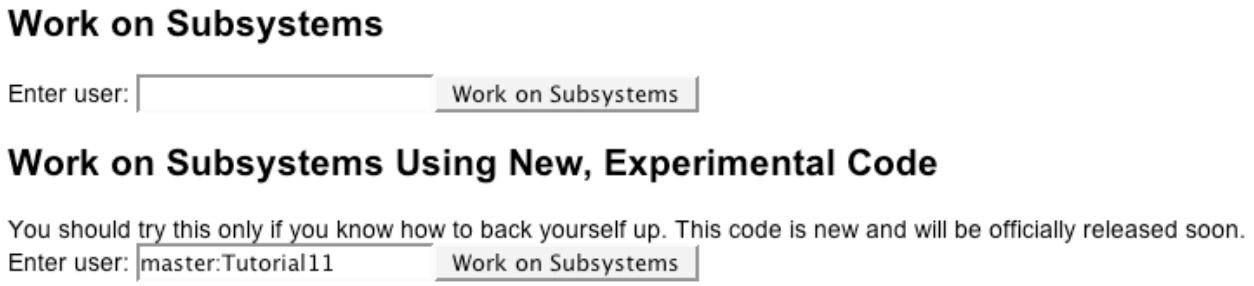


IV.4. NOTES section at the bottom of each SS page contains free-form annotator's comments, lists open problems identified during SS construction and analysis, and - most importantly - explains **variant codes,** which are listed for each organism in a **Variant code** column of a SS spreadsheet . While defining a subsystem, annotators include a collection of functional roles broad enough to cover distinct variations in all relevant organisms. Each subset of functional roles that exists in at least one organism with an operational version of the subsystem constitutes a **functional variant**. Try sorting SS spreadsheet "by_variant".

IV.5. Subsystem diagram (graphic representation of a pathway) is often helpful in analyzing a SS and assigning variant codes. Graphic map of your SS can be accessed from the Assignment page on SEED WIKI, from SEED Forum, or by following this link: http://brucella.uchicago.edu/SubsystemForum/showthread.php?t=81. Open it in a new Tab or Window – you'll need it for the next step.

Part V. EXPAND A SUBSYSTEM

V.1. In order to modify an existing Subsystem or to start a new one you will have to enter SEED under your User ID (as it appears on the Assignment page under your Terminal number). If you haven't yet done so, type your User ID in the " Enter user" window under **Work on Subsystems Using New, Experimental Code** caption and press Work on Subsystems button:



V.2. A list of SS available in your version of SEED will be generated. Scroll down this page to SS named **"NAD and NADP tutorial #"** where # matches the number of your terminal. All these **NAD and NADP tutorial #** SSs are copies of the "**NAD and NADP tutorial"** mother-SS prepared beforehand for this tutorial. Procedure for "cloning" a SS, while changing its ownership will be explained in class.

Note that in addition to "Export full" and "Export assignments" several more functions are now available for you, including "reset", "delete", "publish". Since your User ID matches exactly the User ID it was created with - you are a rightful owner of this SS. Click on the SS name to open it.

V.3. Sort organisms in the spreadsheet by phylogeny. Activate "show clusters".

V.4. Add a new genome to your SS:

-- scroll down the SS page to **Pick Organisms to Extend with** window,
-- locate the genome specified in your assignment, and highlight it
-- put a check mark in the little "fill" box below and press update spreadsheet button
-- after SS is reloaded a new row for this organism will be added to the spreadsheet.

In a new Tab open the SS diagram (http://brucella.uchicago.edu/SubsystemForum/showthread.php?t=81), examine it, compare the added organism to its nearest relatives – try to predict which functional roles are missing from the spreadsheet in this genome.

V.5. Find candidate genes for the missing functional roles. To do so, locate "**show missing with matches**" tool. Check the box near this tool. Copy **Genome ID** for your organism (NOT the organism name) and paste it into "To restrict to a single genome" window near this tool. Press update spreadsheet button. This activates an automatic search for PEGs potentially suitable of filling every empty cell in the spreadsheet row corresponding to the genome specified. In 1 to 8 minutes (depending on the work load on the SEED server) a table with candidate genes will be generated:

To Check Missing Entries:

243277.1: Vibrio cholerae O1 biovar eltor str. N16961 [B]

Assign	P-Sc	PEG	Len	Current fn	Matched peg	Len	Function
☐	9.3e-06	951	397	Ubiquinone biosynthesis monooxgenase UbiF/COQ7 (EC 1.14.13.-)	tr|O15229	486	Kynurenine 3-monooxygenase (EC 1.14.13.
☐	6.7e-06	220	164	Phosphopantetheine adenylyltransferase (EC 2.7.7.3)	sp|P57084	172	Nicotinamide-nucleotide adenylyltransferase 2.7.7.1)
☐	2.8e-07	2947	276	NAD synthetase (EC 6.3.1.5)	fig|155919.1.peg.1818	545	NAD synthetase (EC 6.3.1.5) / Glutamine amidotransferase chain of NAD synthetase

Process assignments

V.7. Carefully examine each candidate by opening the corresponding PEG page (follow the links in PEG column in the table above). It is convenient to open each PEG page in its own Tab or Window in order not to loose the SS page with search results. Analyse the close homologs of the candidate genes, their genome context, annotations of the candidate genes in other databases. Consider functional context as well: using SS diagram rationalize addition of each candidate functional role in the context of NAD(P) metabolism in your specific genome.

V.8. Candidates that you found acceptable need to be re-annotated so that gene functions will match exactly functional roles as they appear in SS. Then candidate PEGs will be automatically connected to your SS spreadsheet.
Candidate genes can be annotated from the Similarities table or Commentary page as described in Part II above. But the simplest way to do this is from the **Missing Entries** table from the SS page (see illustration above):

-- simply mark the check boxes in **Assign** column and press Process assignments button
– selected PEGs will be automatically annotated to match the corresponding functional roles (listed in the right-most column)
-- a report of this transaction will be generated in the same window
-- use Back button on your Browser to return to the SS page
-- put a check mark in the little "fill" box below and press update spreadsheet button to connect candidate PEGs to your spreadsheet

V.9. Compare the NAD(P) biosynthesis in the added organism to that in other species, rationalize, try to assess a possible functional variant. Are there any missing genes or open problems?

If you got that far – CONGRATULATIONS!
You are a certified SEED user!