Example exploratory workflow through the Ocean Protein Portal:

You are interested in understanding how different metals are used by microorganisms in the marine environment. Your favorite metal in Nickel.

- Enter "Nickel" into the search field
- The following screen will show up with proteins that have "nickel" in their descriptions

- You know that *Prochlorococcus* is the most abundant photosynthetic organism on the planet! (fun fact) and so pretty much anything *Prochlorococcus* does is important because of their sheer abundance and contribution to primary production. You select the first *Prochlorococcus* protein that comes up to investigate further.
    - putative nickel-containing superoxide dismutase precursor (NISOD) (EC:1.15.1.1); K00518 superoxide dismutase [EC:1.15.1.1]   NODE_1169163_length_320_cov_1.40377_1_320_+   Prochlorococcus marinus str. MIT 9301

- The first thing you do is see that the protein is a super oxide dismutase. But you're not sure what exactly that means, so you use the tools available to find out more.
    - Click on the"PF09055" link -- it will take you to the Pfam website which tells you that superoxide disumtase is used to clean up free superoxide radicals and convert them to more stable products (O_2, H_2O_2)... This is super important in an oxygenic photosynthesizing organism as these superoxide radicals can cause cell damage if not regulated.
- Now that you know what the nickel superoxide dismutase is, you want to see what the sequence is for this organism. You click on "View Sequence" and it shows you the sequence of the protein. You can click on "blast" button to send this sequence to NCBI blast tools.
- Great! Now you know what this protein is, you want to know where it is in this transect. So you click on the radio button and click "Profile Plot"
    -
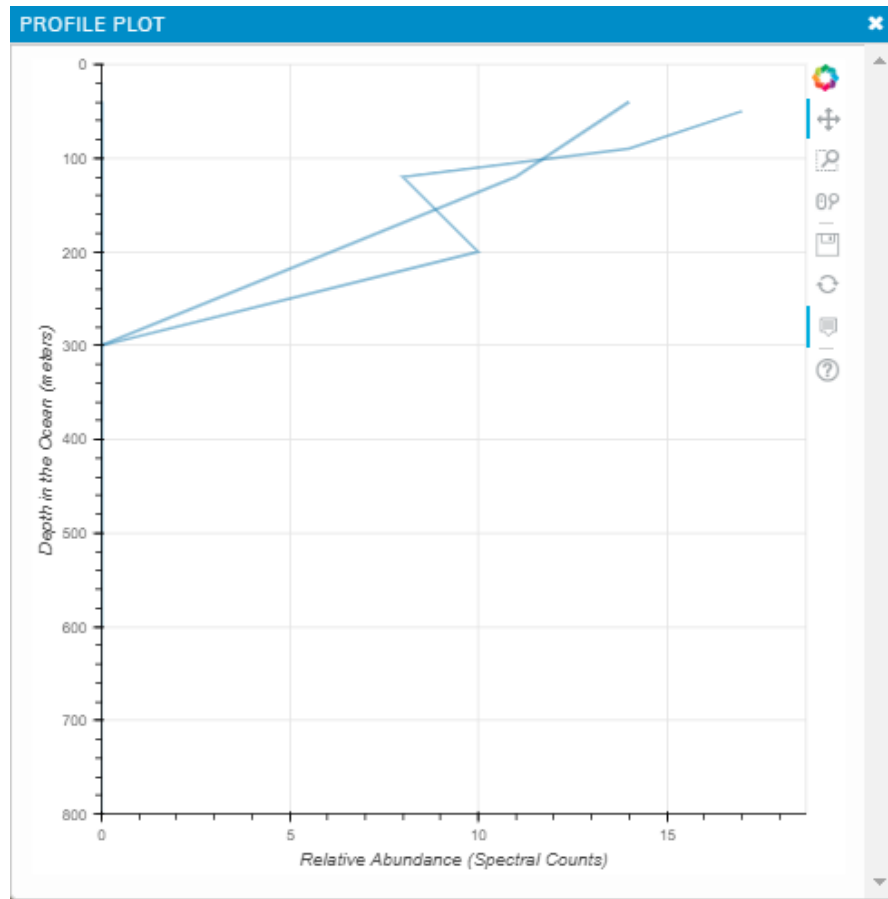
- o
- o   You can see that this protein is more abundant in the surface than at depth
- Click on "View Section"

SECTION HEATMAP FOR NODE_1169163_LENGTH_320_COV_1.40377_1_320_+

Maximum Depth

Number of Contours

Update

o
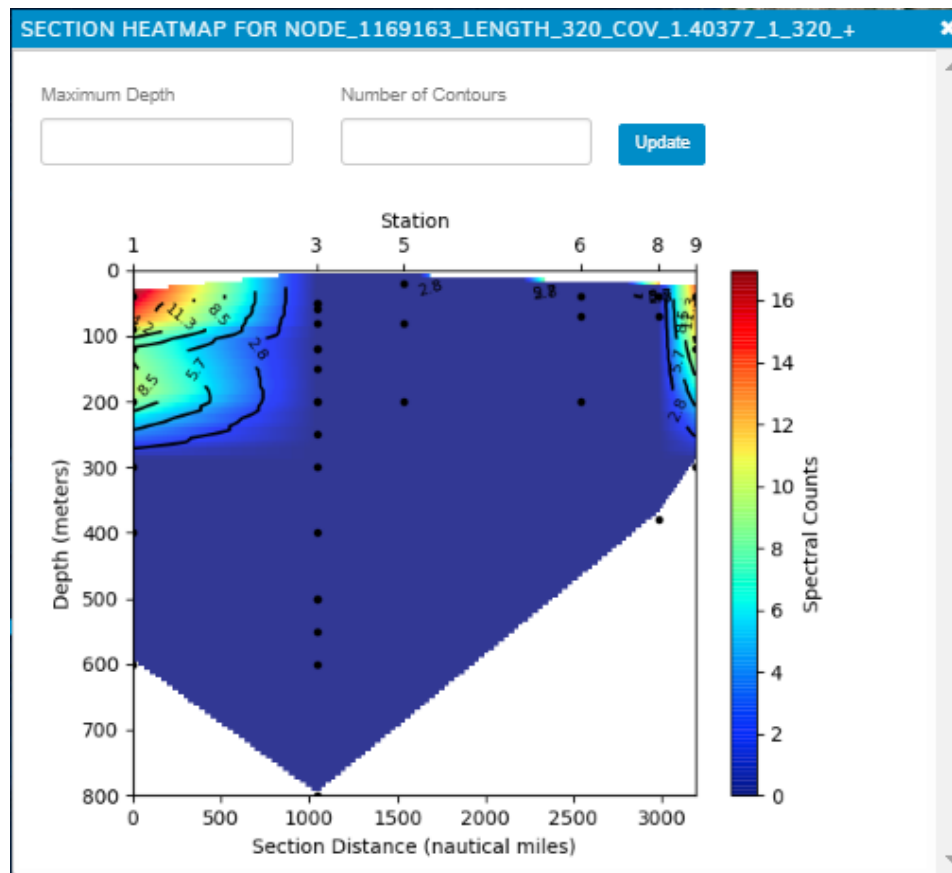
○ Interesting! Appears that this protein is only found in certain locations. Since all the spectral counts are found in the surface, re-focus the section plot on the top 300m **[set max depth to 300 & hit enter]**
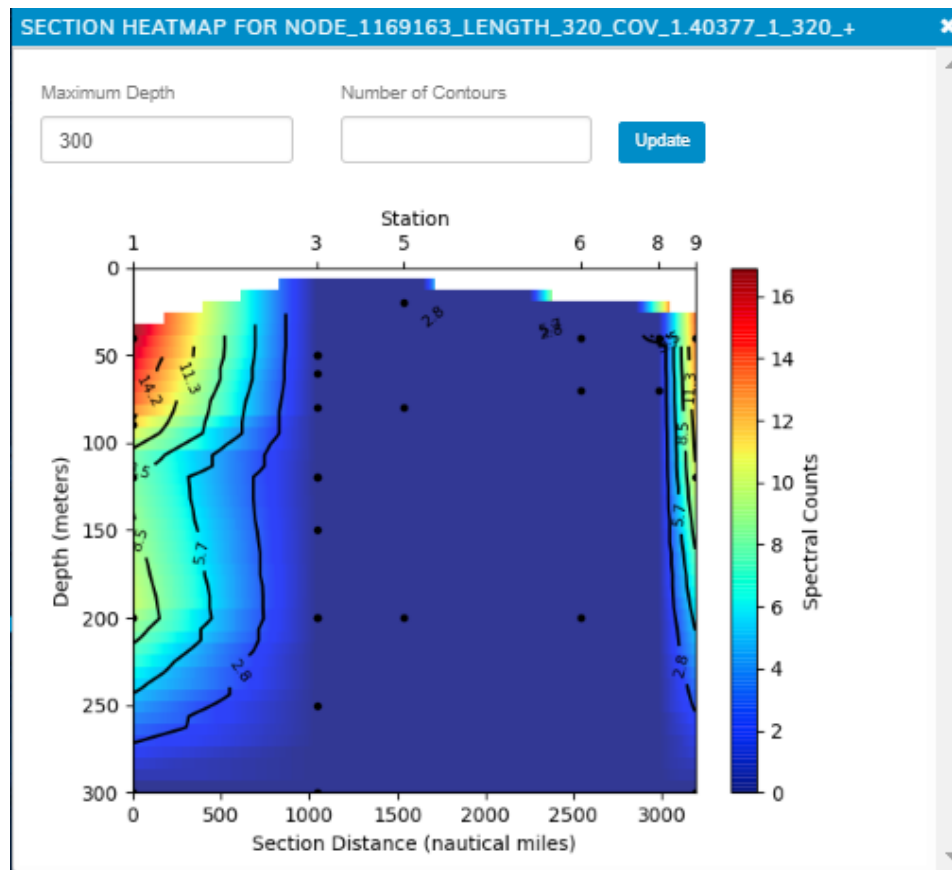
- You know that whole proteins cannot be detected by these analyses... The molecules are just too big for our mass spectrometers! So the proteins are broken up into peptides and it's the peptides which are actually getting measured, with informatics used to estimate protein abundance. You want to look at the peptide information that resulted in this protein information. **[click on "2 peptides" link]** which will redirect you to a new tab "Peptides Found"

- These are the two peptides that were identified from the Nickel Superoxide Dismutase.
- Since peptides are very small... These can be highly conserved and found in many different organisms. You'd like to know what different organisms these peptides may have been from (aside from Prochlorococcus MIT9301 which you saw in the "NCBI Taxon" description of the protein. You click on "Least Common Ancestor" for LIALEAPSSTDSAEWATYSNTFSR.

  -

    LEAST COMMON ANCESTOR FOR LIALEAPSSTDSAEWATYSNTFSR

    Least Common Ancestor is Prochloraceae for peptide ENSLFGVLSLLTGHR

    Bacteria · 1   Cyanobacteria · 1   unclassified Class · 1   Synechococcales · 1   Prochloraceae · 1   Prochlorococcus · 1   Prochlorococcus marinus · 1
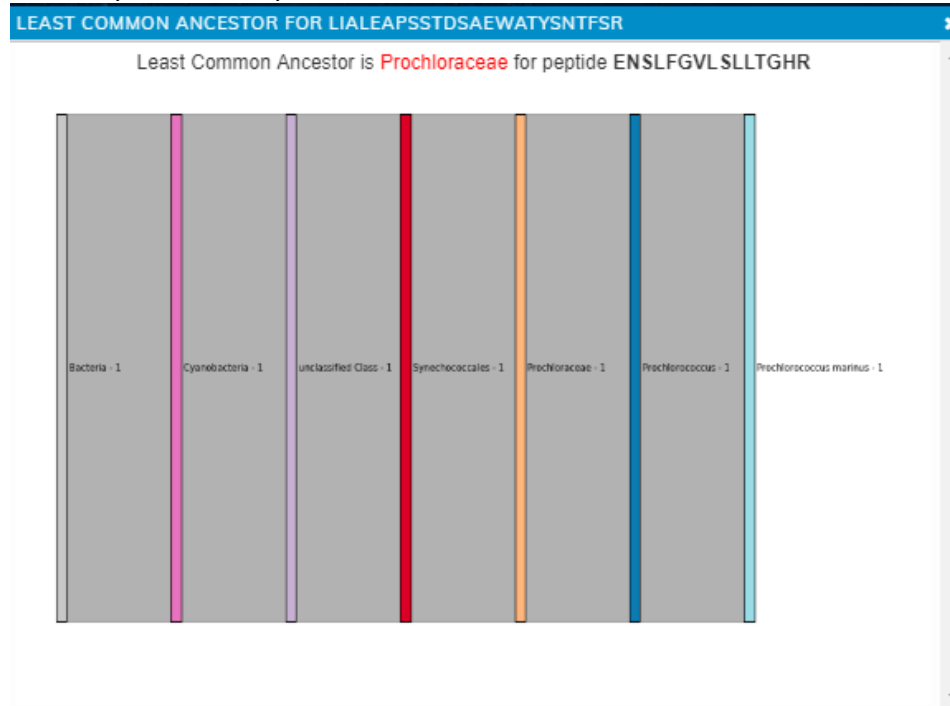
    - You see that this peptide is particular to *Prochlorococcus*. This information is being gathered from METATRYP. METATRYP searches for the peptide against a curated database focused on marine microorganisms. **note: yes, there is a bug in the heading of this Sankey/tree. It is fixed in the portal test server.*
    - You go back and click on the "Least Common Ancestor" link for the other peptide.

LEAST COMMON ANCESTOR FOR VAAEAVLSMTK ✖

Least Common Ancestor is Cyanobacteria for peptide ENSLFGVLSLLTGHR
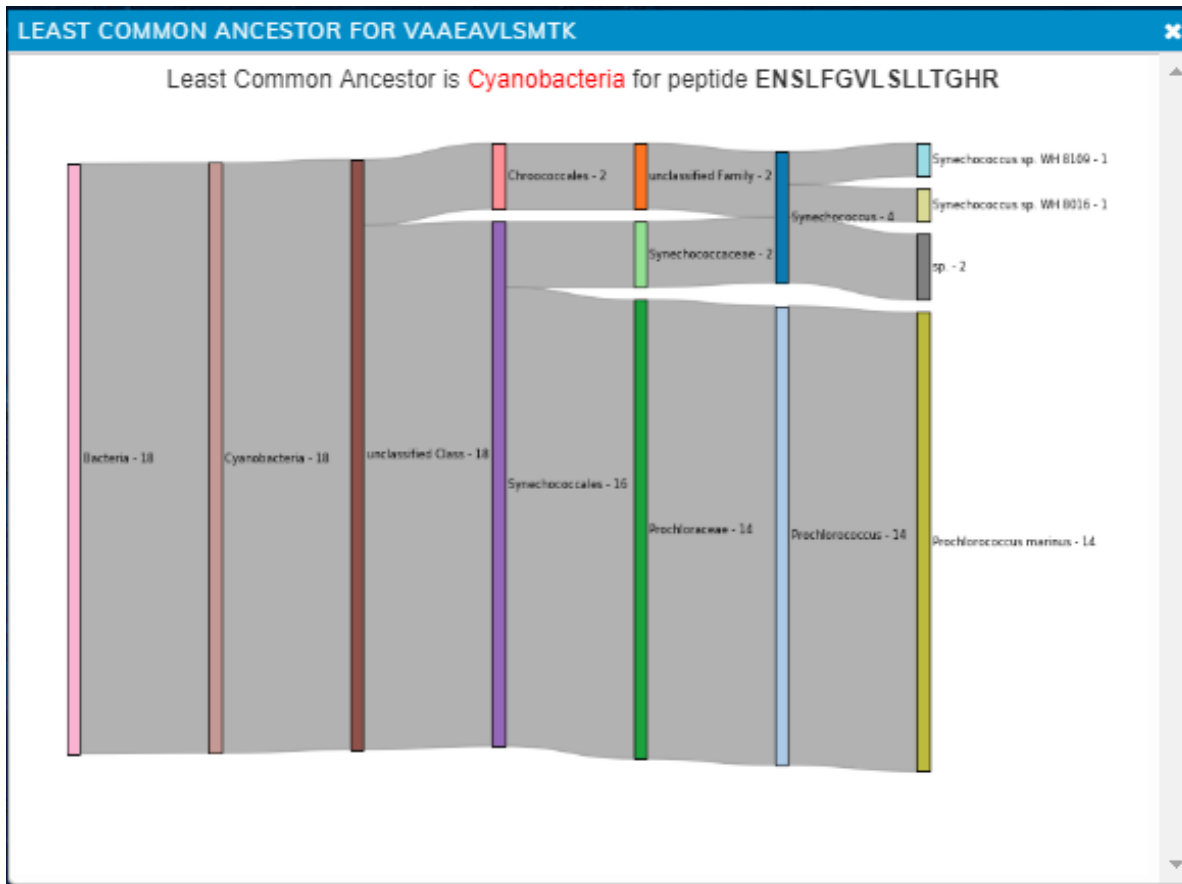
- That's a little more interesting. You see that this is found in more than just *Prochlorococcus*. You want to find more about what these other organisms are, so you explore the METATRYP output more directly. **[Go back to the previous screen and select the hyperlink of the peptide itself VAAEAVLSMTK].**

Least Common Ancestor is Synechococcales for peptide VAAEAVLSMTK

Contig Taxon ID	Contig Taxon	Contig Taxon Level	Annotation	ORF Taxon ID	ORF Taxon	ORF Taxon Level
1218	Prochlorococcus	genus	superoxide dismutase, Ni	scf7180008710917		
1218	Prochlorococcus	genus	superoxide dismutase, Ni	scf7180008710818		
1129	Synechococcus	genus	superoxide dismutase, Ni	scf7180008743347		
1129	Synechococcus	genus	superoxide dismutase, Ni	scf7180008827203		
1129	Synechococcus	genus	superoxide dismutase, Ni	scf7180008884065		
1219	Prochlorococcus marinus	species	superoxide dismutase, Ni	scf7180009003011		
1218	Prochlorococcus	genus	superoxide dismutase, Ni	scf7180009062071		
1218	Prochlorococcus	genus	superoxide dismutase, Ni	scf7180009258987		
1218	Prochlorococcus	genus	putative nickel-containing	scf7180009293166	Prochlorococcus marinus	species

- The METATRYP output for Metagenomes is displayed. What this shows is that from the marine metagenome assemblies in the curated marine microorganism search database, this peptide can be found in these organisms. You know that Synechococcales is the taxonomic group which consists of *Prochlorococcus* and its sister genus *Synechocococcus*

- **[Click on the Peptides Found tab]** Your goal is to look for a specific peptide which can act as a biomarker target for Nickel Superoxide Dismutase in both *Prochlorococcus* and *Synechococcus*. You use the heatmap tool to check which organisms in the METATRYP genome database have these peptides. **[With both peptides selected, click on the Heatmap]**

-
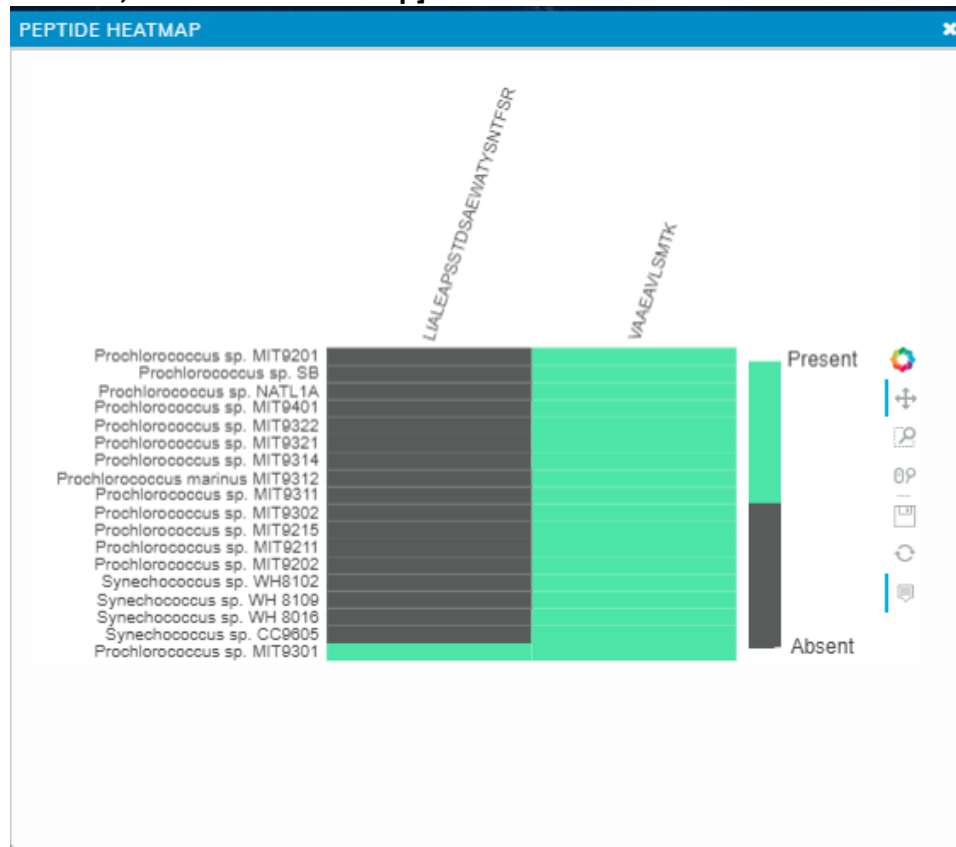
- - Looks like VAAEAVLSMTK would be a much better target for catching Pro & Syn Superoxide dismutase. However, you check the METATRYP database to see just what genomes are in it: https://metatryp.whoi.edu/database/
  - You could then use this information to make a peptide standard of VAAEAVLSMTK and directly quantify how much of this protein is in a sample with targeted proteomics (if you are a proteomics researcher).