**Align sequences using Clustal Omega**

1. Use either UniProt or Saccaromyces Genome Database (for *Saccaromyces cerevisiae*) to download sequences. If using UniProt, proceed to step 2. If using SGD, proceed to step 6.

UniProt

1. Go to uniprot.org. Search gene/protein name or other relevant information in the search bar at the top of the webpage.
2. Click the entry button of the desired entry. This page shows you information about the protein as well as the organism in which it is found.
3. Click the “BLAST” button just under the entry name. Use the default setting.
4. The protein sequence should now appear. Copy this sequence to your clipboard.
5. To more easily search all UniProt results for your desired organism, you can download all results as a FASTA (canonical) file and open using Excel, then search for your organism.

Saccaromyces Genome Database

1. Go to yeastgenome.org. Search gene/protein name in the search bar in the top right of the webpage. If this does not pull up information about your protein, you may need to select a protein from the drop down menu of the search bar.
2. Click the “Download” button in the Sequence section. Select whichever type of download applies to you (“Protein” for sequence alignment).
3. Open this download with any file that allows you to see the sequence (I typically use Microsoft Word). Copy this sequence to your clipboard.

Clustal Omega

1. Go to <https://www.ebi.ac.uk/Tools/msa/clustalo/>. This should open the input form for a multiple sequence alignment.
2. Paste your copied sequences into the “sequences in any supported format” box. In the line above the start of your sequence, there should be a ‘greater than’ sign followed by whatever you would like this sequence to be called (>S\_cerevisiae)(i.e., FASTA formatted). Note that spaces cannot be used. Sequences can be pasted sequentially, in the same text box. Click submit.
3. Your sequence alignment should now be visible. Go to Results Viewer and copy the url under “View result with Jalview.”

**Analyze and Save Sequence Alignment in JalView**

Jalview

1. Open Jalview on your computer. Close all the example windows that may pop up.
2. Click “File,” then “Input Alignment,” then “from URL.” Paste the URL from Clustal Omega. Your alignment should now open in Jalview. Save this file.

**Analyze Protein Structure and Map Conservation with UCSF Chimera**

1. You will first need a protein structure. Go to the Protein Data Bank at rcsb.org. Search for desired protein in the search bar.
2. You can click on the name of the structure for more information about it, including the paper it’s originally from. Take note of the four-character ID code.
3. Open Chimera. Click “Fetch” in the bottom right corner of the window.
4. In the new window that opens, select PDB and type the four-character code from step 2. Type enter, and the structure from PDB should now be visible.

Using Chimera (MacBook)

1. Click and move one finger on the track pad to rotate the structure.
2. Click with two fingers to zoom in and out.
3. Click with three fingers to move the structure without rotating.
4. Press Control + click on a residue to select it. If you have an alignment open, this will highlight the residue in the alignment.
5. Control+shift+click to deselect a residue.
6. Select a residue, click Actions, Atoms/Bonds, side chain/base, then show to display side chains.
7. Select a residue, click Actions, then Colour, to color residues. This is useful for mapping mutations.
8. Click Tools, General Controls, then Command Line. This allows you to easily search for residues (ex. select 149.a will select residue 149 in chain A).

Mapping by Conservation

1. You may want to map by conservation to visualize how conserved each residue is across species. To begin, you must open your sequence alignment from Jalview. Click File, then Open, and select your Jalview alignment from your computer.
2. Determine which chain of the protein you want to map by hovering your cursor over the structure and looking at the chain names in the white bar at the bottom of the window.
3. In the alignment window, click Structure, then Associations. You want to associate the chain you’re analyzing with a certain species from your alignment (usually the one from which the structure was derived, for analyses in Matt’s lab it will typically be budding yeast or human). Click OK.
4. Click Preferences, then Headers. Make sure the Conservation Style is set to AL2CO. Click OK.
5. Click Structure, then Render by Conservation. Set the attribute to mavConservation. Click OK.
6. Your chain should now be mapped according to conservation. Red is more conserved, blue is less conserved.

**Faster option for fetching and aligning sequences entirely in Jalview:**

1. Open Jalview
2. File > Fetch sequences> select database – Uniprot > enter gene of interest
3. Select multiple entries at same time by holding down command (on mac) and selecting desired entry
4. Check that you are selecting the right protein and not a similar protein. For example, a search for Spc19 will also bring up some entries for Spc34 and other Dam1c components
5. Hit enter to bring them into your alignment window
6. Can reorder by dragging the entries
7. Select all entry names > Web service > alignment > Clustal (or whatever method you prefer) > use default parameters (or whatever parameters you want)
8. To color: we usually color by percent identity. Colour > percentage identity

Note: This method won’t always have all the species you want. You can explore the other databases under the fetch sequences menu to find other species that aren’t in uniprot. I find uniprot is usually the best place to start though and sufficient for a quick look.

Emily’s notes –

* I think it would be helpful to have 2 ways of doing the alignment for fungal species – one using a species every 10-20 that are coming up when blasting on uniport and one from a list of “typical” strains to use (in which one of us looks up a general phylogenetic tree for fungal species & “intelligently” selects species that represent various degrees of divergence). Maybe it would make sense to do the same for alignments that go up to human.
* When I do the blasts, I only get 100 sequences; is there a way to get more than this?
* Depending on the protein, I can also get proteins with a different name, and this might make automation tricky. For example, Spc105 has many blast hits that are “Spc7” in those organisms, and some are just “Spc7-domain containing protein”. I don’t know whether this will be an issue or not.