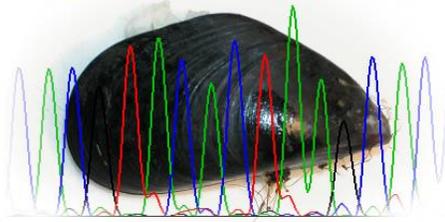


C A C G T A C A C T A A G C A C



Marine Biotechnology and Bioinformatics for Teachers

MARINE BIOTECHNOLOGY & BIOINFORMATICS FOR TEACHERS MOSS LANDING MARINE LABS NSF ITEST GRANT STUDENT WORK SHEET USING MOLECULAR HOMOLOGIES TO PREDICT EVOLUTIONARY RELATIONSHIPS

BACKGROUND

Proteins are amazing! Without them, most reactions in our bodies would not occur at a speed fast enough to sustain life. They are the workhorse of the cell and they are ultimately what cause each of us to function and be unique.

Even though living organisms are so diverse, some proteins have been conserved over evolutionary time. This means that the protein that breaks down waste in your cells is also breaking down waste in the cells of your dog, your favorite plant, or the yeast in your bread. Based on the similarities in these protein sequences, scientists can estimate the relative relatedness between two organisms.

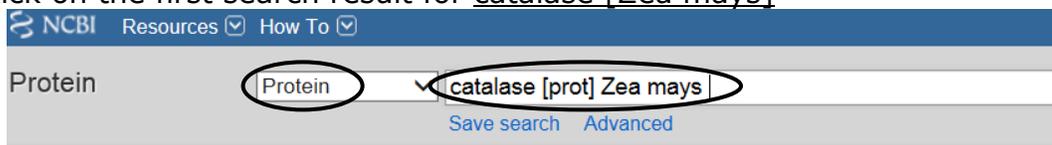
Today, you will be using some common research tools to analyze the protein sequence of a common protein, catalase, in several organisms. You will then compare these protein sequences to look for similarities and evidence of evolutionary relationships. In this web activity, you will be using bioinformatics to answer three main questions:

- Do proteins with the same function in different organisms have the same structure and amino acid sequence?
- What can protein structure and sequence tell us about evolutionary relationships?

PROCEDURE

Part 1: Do proteins with the same function in different organisms have the same structure and amino acid sequence?

1. Go to the NCBI website at <http://www.ncbi.nlm.nih.gov/>
2. Select "Protein" from the search drop down (All Databases) menu and type "catalase [prot] Zea mays" into the search window. Click **Search**
This will give you a list of catalase protein sequences from various organisms.
3. Click on the first search result for catalase [Zea mays]



[Show additional filters](#)

Display Settings: Summary, 20 per page, Sorted by Default order

Species

Plants
Bacteria
More ...

Source databases

EMBL
GenBank
RefSeq

Items: 8

[catalase \[Zea mays\]](#)

1. 496 aa protein

Accession: AAC07387.1 GI: 168437

[GenPept](#) [FASTA](#) [Graphics](#) [Related Sequences](#) [Identical Proteins](#)

[catalase \[Zea mays\]](#)

2. 491 aa protein

Accession: AAC07440.4 GI: 168438

- Select "FASTA." (under Zea Mays that you checked in step 3)
This will give you the protein sequence in single-letter amino acid code.
- Copy the protein sequence starting from ">..." and paste this FASTA format sequence into a new document, for example Notepad. (On a PC, notepad is under Start → Programs → Accessories → Notepad)

```

Untitled - Notepad
File Edit Format View Help
>AAC37357.1 catalase [Zea mays]
MTMDPTKFRPSSSHDITVTTNAGAPVWNDNEALTVGPRGPILLEDYHLIEKVAHFARERIPERVVHARG
ASAKGFFECHDVTSLTCADFLRAPGVRTPVIVRFSTVIHERGSPETIRDPRGFVAVKFYTREGNWDLLGN
NFPVFFIRDGIKFPDVIHAFKPNPRSHVQYWRVDFLHSLPESLHTFFFLLFDDVGVPSDYRHMEGFGVN
TYTFVSAAGKAQYVKFHWKPTCGVRCILTDEEALVGGRNHSHATQDLYDSIAAGSFPEWTLVQVMDPD
TEEQYDFDPLDDTKTWPEDLLPLRPVGRVLVLDNRVDFNFFNENEQLAFGPGLVVPGIYYSDDKMLQCRVFA

```

- Search the web to find the common name for this organism and add this name in place of the numbers and letters before the term catalase and [Zea mays] into your notepad document. ** You must keep the '>' before the scientific name and common name –see example below.

```

>corn catalase [Zea mays]
MTMDPTKFRPSSSHDITVTTNAGAPVWNDNEALTVGPRGPILLEDYHLIEKVAHFARERIPERVVHARG
ASAKGFFECHDVTSLTCADFLRAPGVRTPVIVRFSTVIHERGSPETIRDPRGFVAVKFYTREGNWDLLGN
NFPVFFIRDGIKFPDVIHAFKPNPRSHVQYWRVDFLHSLPELHTFFFLLFDDVGVPSDYRHMEGFGVN
TYTFVSAAGKAQYVKFHWKPTCGVRCILTDEEALVGGRNHSHATQDLYDSIAAGSFPEWTLVQVMDPD

```

- Then, save the untitled.txt notepad file as "catalase_protein_seqs.txt" on your desktop.
- Return to your original NCBI search "catalase [prot]." and replace Zea mays with the organism *Prunus persica* Repeat Steps 3-6 making sure to paste the next protein into the same "catalase_protein_seqs.txt" file. Do this for AT LEAST SIX organisms total. (If you have chosen more than one type of bacteria give them the names bacteria1 and bacteria2 (omit spaces in your common name)

```

>corn catalase [Zea mays]
MTMDPTKFRPSSSHDITVTTNAGAPVWNDNEALTVGPRGPILLEDYHLIEKVAHFARERIPERVVHARG
ASAKGFFECHDVTSLTCADFLRAPGVRTPVIVRFSTVIHERGSPETIRDPRGFVAVKFYTREGNWDLLGN
NFPVFFIRDGIKFPDVIHAFKPNPRSHVQYWRVDFLHSLPELHTFFFLLFDDVGVPSDYRHMEGFGVN
TYTFVSAAGKAQYVKFHWKPTCGVRCILTDEEALVGGRNHSHATQDLYDSIAAGSFPEWTLVQVMDPD
TEEQYDFDPLDDTKTWPEDLLPLRPVGRVLVLDNRVDFNFFNENEQLAFGPGLVVPGIYYSDDKMLQCRVFA
YADTQRYRLGPNYLMPLPVNAPRCAHHNNHHDGAMNFMHRDEEVYYPSPRHAPLRQAAPPTPLPPRFVAGR
REKATIRKPNDFKQPGERYRSWADPDRQERFIHRWVDALSDPRVTHEIRSVWISYWSQADKSLGQKLASRLNVR
PSI

```

```

>peach catalase [Prunus persica]
SAKGFFEVDHISQLTCADFLRAPGVQTPVIVRFSTVIHERGSPETLRDPRGFVAVKFYTREGNFDLVGNN
FPVFFVRDAMKFPDAIRAFKPNPKSHIQETWRILDFFSHLPESLHTFAFFYDDLGVPPQDYRHMEGSSVHA
YTLISKAGKVHYVKFHWKPTCGVKCLLEDEAIKVGGANHSHATKDLYDSIAAGNYPEWKLYIQTMDDPDHE
DRFDFDPLDLTKTWPEDILPLQPVGRVLVKNIDNFFAENEQLAFNPAHVVPYIYSDDKMLQTRIFAYS
DTQRHRLGPNYLQLPVNAPKCPHHNNHHDGAMNFMHRDEEVYYPSPRHAPLRQAAPPTPLPPRFVAGR
CVIEKENNFKQPGERYRSWADPDRQERFIHRWVDALSDPRVTHEIRSVWISYWSQADKSLGQKLASRLNVR
PSI

```

```

>human catalase [Homo sapiens]
MADSRDPASDQMQRHWKEQRAAQKADVLTTGAGNPVGDKNVITVGPGRGPLLVQDVVFTDEMAHFRERIP
ERVVHAKGAGAFGYFEVTHDITKYSKAKVFEHIGKKTPIAVRFSTVAGESGSADTVRDRPRGFVAVKFYTED

```

- Once you have compiled catalase sequences for at least six organisms, go to the Clustal website from Kyoto University Genome Net at <http://www.genome.jp/tools/clustalw/>

10. Copy and paste all six sequences from the Notepad window into the Clustal window.

Clustal is a program that aligns the amino acid sequences based on their similarities. (Each amino acid is abbreviated with one letter of the alphabet.)



Multiple Sequence Alignment by CLUSTALW

ETE3 MAFFT **CLUSTALW** PRRN

General Setting Parameters:
 Output Format:
 Pairwise Alignment: FAST/APPROXIMATE SLOW/ACCURATE

Enter your sequences (with labels) below (copy & paste): PROTEIN DNA

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

```
>corn catalase [Zea mays]
MTMDPTKFRPSSSHDITVTTNAGAPVWVNDNEALTVGPRGPILLEDDYHLIEKVAHFARER
IPERVVHARG
```

```
catalase_protein_seqs - Notepad
Edit Format View Help

prn catalase [Zea mays]
MDPTKFRPSSSHDITVTTNAGAPVWVNDNEALTVGPRGPILLEDDYHLIEKVAHFARERIPERVVHARG
KGFECTHDVTSITCADFLRAGVVRTPVIVRFSTVIHERGSPETLRDPRGFVAVKFTYREGNFDLVGN
PVFFTRDGIKFPDVIHAFKPNRSHVQYWRVDFLHLPESLHTFFFLDQVGVPSDYRHMGEFGVNI
TFVSAAGKAQYVKFHKPTCGVRCILTDEEALVGGRNHSHATQDLYDSIAAGSFPEWTLVQVMDPD
EQYDFDPLDDTKTWPEDLPLRPVGRVLVDRNVDNFFNEIEQLAFGPGLVWPGIYSDDKMLQCRVFA
DTQRYRLGPNYMLPVNAPRCAHHNNHYDGAMNFMHRDEEVDYYPSPRHAPLRQAAPPTLPPRPVAGR
KATIRKPNDFKQPGERYRSDADRQDRVFRFADSLGHPKVSQELRSIWDLLAKCDASLGMKIATRL
KANM

each catalase [Prunus persica]
PYKRRPSSAFDPSYWTNAGAPVWVNDSSLTVGPGRGPVLLDYHLVEKLFATDRERIPERVVHARGAS
GFFEVTHDISQLTCADFLRAGVQTPVIVRFSTVIHERGSPETLRDPRGFVAVKFTYREGNFDLVGN
FEVDRAMKEDDRAITRAEKPNPKSHTOETWRTLDEESHLPESLHTEAEVDDLQVDDVYRHMGEFSSVHAV
```

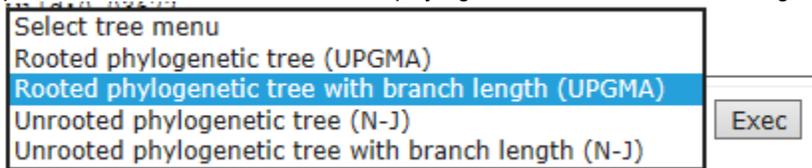
11. Click "Execute Multiple Alignment." Scroll down to see how Clustal aligns the sequences.

Part 2: What can protein structure tell us about evolutionary relationships?

When you look back at your Clustal alignment, you will notice that the program has rearranged the order of the sequences and grouped them together based on their similarities. Clustal also compares the aligned sequences and measures how different they are from each other. The more differences, the less related they should be, and the more distant they are in terms of evolutionary relationships.

*What do stars * mean?*

1. Scroll down to the bottom of the "CLUSTALW results" web page.
2. Click on the drop-down menu and select "Rooted phylogenetic tree with branch length (UPGMA)."



3. Click "Exec" and a tree that shows evolutionary relationships should appear in a different window.
4. Copy and paste your phylogenetic tree onto your document, under the amino acid sequences.

EVIDENCE:

1. What can you conclude based on your evolutionary tree?
2. Do the connections make sense? Are any of the relationships surprising to you?

Congratulations! You've navigated through many complicated websites that researchers often use, hope you found this experience with bioinformatics to be interesting! :o)