**Investigating Sequence Similarity**

**Exercise 4.1: Evolution of Alcohol Metabolism**

**Instructor Preface**

This document includes:

1. A student handout for Exercise 4.1
2. A postlab worksheet containing a compilation of the questions integrated throughout the student handout for Exercise 4.1 that students can complete and turn in for a post-lab assignment
3. An appendix with the FASTA files required for this project

Instructors should distribute the student handout and the question compilation worksheet to students, and make a separate FASTA file with the sequences to distribute as a separate document.

**Investigating Sequence Similarity**

EXERCISE 4.1

**Inquiry-Based Investigation – Evolution of Alcohol Metabolism**

**Objectives**

After completing this exercise, you should be able to:

1. Apply bioinformatics tools to address a real-life biological question

How did humans evolve the ability to metabolize alcohol? Did it arise from (i) the fact that early humans stored large quantities of food and inadvertently fermented them, or did it arise in response to (ii) primates moving from an arboreal (tree-living) to a terrestrial lifestyle, where fermented (i.e., rotting) fruits on the ground are in abundance? Your investigation will parallel the efforts of evolutionary biologists seeking to understand how alcohol metabolism evolved in hominids (Carrigan et al., 2015).

To investigate this question, use your knowledge of bioinformatics tools to retrieve FASTA sequences of the alcohol dehydrogenase enzyme from **Genbank**, **ADH4**, perform a multiple sequence alignment, and generate a phylogenetic tree in **MEGA**.

Answer Postlab Question 1 at the end of the document

1) Start by retrieving the FASTA protein sequences using the NCBI accession numbers in **Table 1** and paste them into the instructor provided .fas file. You will notice that a subpopulation of the sequences, which are housed in a separate database, have already been imported into the .fas file to speed up your analysis.

* You can search by **accession number** in Genbank (you will have to find the website, we’ve done in before) to get the protein sequence.
* Once you’ve found the correct record (the species name should match the species name from the table), click on the FASTA link (see below) to get to a page with the sequence in FASTA format. Then, you should easily be able to cut and paste to add this new sequence to your existing .fas file (we’ve downloaded a few for you). **Remember to do this in a proper text editing program, i.e. TextMate for Mac or NotePad for PC, to maintain it in .fas format.** While you are in your text-editing program, be sure to re-name the sequence to make it easier to read (I’ve been using the common name, then Genbank or Ensemble #, then species name)



2) Next, using MEGA, align the ADH4 protein sequences from arboreal monkeys and terrestrial hominids and create a phylogeny to reveal sequence similarity and infer phylogeny.

* We have been working most recently on **nucleotide alignments—**remember, when MEGA asks you, to choose a **protein** sequence alignment. If you need help doing this, check back on our first phylogeny lab
* Remember, after doing the alignment, export as a .MEG file to do the tree
* I suggest doing a standard **maximum likelihood tree** with ~500 bootstrap replicates
* Make sure the tree has **branch lengths** (all of the tips won’t line up perfectly), **Image>Save as pdf,** and save a copy of the tree to use in your postlab.

**Table 1**. Hominid habitat type and ADH4 protein accession numbers for the National Center for Biotechnology Information and Ensemble databases (Ensemble is a genome database that we haven’t used yet—those sequences will be provided in the .fas file)

| **Hominid** | **NCBI (GenBank) or Ensemble Accession Number** | **Habitat** |
| --- | --- | --- |
| Tarsier (*Tarsius syrichta*) | ENSTSYG00000004598 | Arboreal |
| Marmoset (*Callithrix jacchus*) | AIY30143 | Arboreal |
| Gibbon (*Nomascus leucogenys*) | XP\_003257551 | Arboreal |
| Orangutan (*Pogo pygmaeus*) | ENSPPYT00000017372 | Arboreal |
| Bonobo (*Pan paniscus*) | XP\_003829978 | Terrestrial and Arboreal |
| Chimpanzee (*Pan troglodytes*) | XP\_016807404 | Terrestrial and Arboreal |
| Gorilla (*Gorilla gorilla*) | ENSGGOG00000011864 | Terrestrial |
| Human (*Homo sapiens*) | NP\_000664 | Terrestrial |

*Answer Postlab question 2 at the end of the document*

In the developing field of paleogenetics, phylogenetic analysis of protein sequences can allow biologists to resurrect ancestral proteins from extinct organisms and test their functionality in the laboratory. This experimental approach can yield evidence about when a gene gained a novel or differential level of functionality through evolutionary time and provide context for identifying evolutionary pressures that shaped these events.

If the arboreal and terrestrial hominid protein sequences cluster together, is there a correlation with hominids moving from trees to a terrestrial lifestyle and enzymatic activity of an ancestral ADH4 enzyme from each of these clades (**Table 2**)?

**Table 2**. Enzyme Efficiency of Resurrected Ancestral ADH4 Enzymes with Ethanol Substrate for Selected Clades of Hominids.

|  |  |
| --- | --- |
| **Hominid Clades** | **ADH4 Enzymatic Efficiency (mM−1•min−1)** |
| Human, Bonobo, Chimpanzee, Gorilla | 32 ± 6.2 |
| Orangutan (Pogo pygmaeus) | 0.87 ± 0.17 |
| Gibbon (Nomascus leucogenys) | 0.45 ± 0.05 |
| Baboon, Macaque, vervet | 0.57 ± 0.14 |
| Marmoset, Tamarin, Capuchin, Squirrel Monkey | 0.69 ± 0.04 |
| Tarsier | 0.45 ± 0.06 |

*Answer postlab question 3. Do the data support one of the proposed hypotheses? Explain, clearly referring to some of the specific values in Table 2.*

**References**

Carrigan, M. A., Uryasev, O., Frye, C. B., Eckman, B. L., Myers, C. R., Hurley, T. D., & Benner, S. A. (2015). Hominids adapted to metabolize ethanol long before human-directed fermentation. *Proceedings of the National Academy of Sciences*. 112(2):458-63. doi: 10.1073/pnas.1404167111.

**Investigating Sequence Similarity**

**Postlab**

**Exercise 4.1 – Inquiry-Based Investigation – Evolution of Alcohol Metabolism**

Postlab question 1: The intro proposed 2 hypotheses (4 pts):

1. Ability to metabolize alcohol evolved in humans who grew and stored food
2. Ability to metabolize alcohol evolved in terrestrial primates who ate inadvertently fermented (boozy) fruit

What would you expect a phylogeny of the ADH genes to look like under hypothesis 1 vs hypothesis 2? (Remember, here the genes are not neutrally evolving, so genes with similar functions should group together, regardless of the evolutionary history of the group).

Postlab question 2: What does your tree look like? Do the arboreal and terrestrial primate protein sequences cluster together? Upload a pdf of your tree and refer to it in your answer, **annotating on the tree** which species are terrestrial and which are arboreal (3 pts)

P*ostlab question 3. Do the data support one of the proposed hypotheses? Explain, clearly referring to some of the specific values in Table 2 (3 pts)*

**Appendix 1: FASTA-formatted sequences for this exercise**

>Tarsier\_ADH.4\_Tarsius\_syrichta\_ENSTSYG00000004598

MGTAGKVIKCKAAVLWELKKPFSIEEIEVAPPKVKEVRIKILATGICRTDDHVIKGAMVS

KFPVIVGHEAAGVVESIGEGVTTVKPGDKVIPLFLPQCRECNACRNPDGNLCIRSDVTGR

GVLADGTTRFTCKGKPVYHFMNTSTFTEYTVVDESSVAKIDDSAPPEKVCLIGCGFSTGY

GAAIKTGKVKPGSTCVVFGLGGVGLSVIMGCKSAGASRIIGIDLNKDKFEKAMAVGATDC

ISPKDYTKPISEVLSEMTGNTVGYSFEVIGRLDTMIDALASCHMNYGTSVVVGAPPSAKM

LTYDPMLLFTGRTWKGCVFGGWKSRDEVPKLVTDFLAKKFDLDQLITHVLPFKKINEGFE

LLYSGQXXXXXXXX

>Marmoset\_ADH.4\_Callithrix\_jacchus\_AIY30143

MGTAGKVIKCKAAVLWDQKKPFSIEEIEVAPPKTKEVRIKILATGICRTDDHVIKGTMVS

KFPVIVGHEATGIVESIGEGVTTVKPGDKVIPLFLPQCRECNACRNPDGNLCIRSDITGR

GVLADGTTRFTCKGKPVYHFMNTSTFTEYTVVDESSVAKIDDTAPPEKVCLIGCGFSTGY

GAAVKTGKVKPGSTCVVFGLGGVGLSVIMGCKSAGASRIIGIDLNKDKFEKAMAVGATEC

ISPKDSTKPISEVLSEMTDNNVGYTFEVIGRLDTMIDALASCHMNYGTSVVVGAPPSAKM

LTYDPMLLITGRTWKGCVFGGLKSRDDVPKLVTEFLAKKFDLDQLITHVLPFKQINEGFE

LLNSGQSIRTVLTF

>Orangutan\_ADH.4\_Pongo\_pygmaeus\_ENSPPYT00000017372

MGTAGKVIKCKAAVLWEQKQPFSIEEIEVAPPKTKEVRIKILATGICRTDDHVIKGTMVS

KFPVIVGHEATGIVESIGEGVTTVKPGDKVIPLFLPQCRECNACRNPDGNLCIRSDITGR

GVLADGTTRFTCKGKPVHHFMNTSTFTEYTVVDESSVAKIDDAAPPEKVCLIGCGFSTGY

GAAVKTGKVKPGSTCVVFGLGGVGLSVIMGCKSAGASRIIGIDLNKDKFEKAMAVGATEC

ISPNDSTKPISEVLSEMTGNNVGYTFEVIGHLETMIDALASCHMNYGTSVVVGAPPSAKM

LTYDPMLLFTGRTWKGCVFGGLKSRDDVPKLVTEFLAKKFDLDQLITHVLPFKKISEGFE

LLNSGQSIRTVLTF

>GOrilla\_ADH.4\_Gorilla\_gorilla\_ENSGGOG00000011864

MGTAGKVIKCKAAVLWEQKQPFSIEEIEVAPPKTKEVRIKILATGICRTDDHVIKGTMVS

KFPVIVGHEATGIVESIGEGVTTVKPGDKVIPLFLPQCRECNACRNPDGNLCIRSDITGR

GVLADGTTRFTCKGKPVHHFMNTSTFTEYTVVDESSVAKIDDAAPPEKVCLIGCGFSTGY

GAAVKTGKVKPGSTCVVFGLGGVGLSVIMGCKSAGASRIIGIDLNKDKFEKAMAVGATEC

ISPKDSTKPISEVLSEMTGNNVGYTFEVIGHLETMIDALASCHMNYGTSVVVGVPPSAKM

LTYDPMLLFTGRTWKGCVFGGLKSRDDVPKLVTEFLAKKFDLDQLITHVLPFKKISEGFE

LLNSGQXXXXXXXX