Comparing Amino Acid Sequences

Abstract

In this hands-on activity, students work in small groups to compare amino acid sequences for a particular protein from a mouse to the same protein from 5 other organisms, then calculate the percent of amino acids that are identical. All groups combine their data, revealing a ***pattern*** in the percent identity to mouse for 10 organisms across 5 different proteins in all.

Learning Objectives

- Vastly different organisms make similar proteins.
- A core set of genes is required for basic life functions; these functions are common to all types/domains of organisms.
- There is a greater percent identity in proteins among organisms that are more simlar to each other.

Estimated Time

- 45-60 Minutes

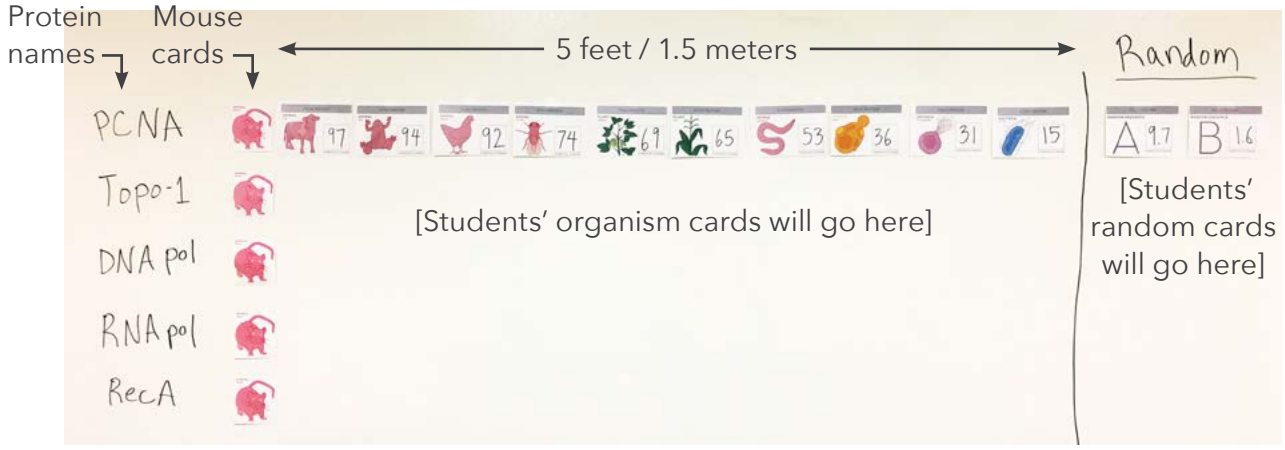
Materials

- Scissors, tape, and calculators; markers are optional
- Copies of Student Handout (pdf); follow instructions in that document.
- White board or other large blank space to arrange data cards
- Mouse cards (from page 4 of this document)

> ***A Note on Materials:*** *Use of color makes the patterns of similarities and differences stand out more when the organism cards are placed in their final arrangement.*
>
> *If printing in color is a barrier, consider printing just one set of the organism cards in color (available in a separate PDF file) and laminating them. Students can write on them with dry-erase markers, and cards may be re-used.*
>
> *If printing in color is impossible, you may use markers or highlighters to color over black-and-white copies, and then laminate them for re-use.*
>
> *Note that if you choose to make a re-usable classroom set of organism cards, another PDF file is available with just the amino acid sequences (e.g., no cards).*

Instructions

1. Set up whiteboard or wall space with gene names, mouse cards, and Random label.



2. Divide students into 10 groups. Distribute copies according to instructions in the Student Handouts document.

3. Instruct students to write the % identical value in large, bold numbers so that they can be seen from a distance. Use a marker for best results.

4. When all groups have completed their work, have students arrange their data cards on the wall or white board to the right of the matching label. Tell them to arrange their cards in order of % identical: highest to the left, lowest to the right, and random to the right of the vertical line (see key on page 5). The data for each gene is split between two groups (6 cards per group), so you may want to ask a representative from each group to get together and collate their cards before taping them to the board. There is some overlap.

5. You can check students' work using the amino acid alignments included in this document.

Note: This activity uses standard one-letter abbreviations for amino acids. The same abbreviations are used in Paper Transcription and Translation; if students need a visual reminder, show them the Amino Acid Codon chart in the Student Instructions for that activity.

Wrap up

You may want to remind your students of the following:

- Some proteins are made by all living things. For these proteins, the percent of amino acids that are identical is generally greater between organisms that are more closely related.
- The data show how closely related all of these sequences are only to mouse. Since all of the comparisons were made to the mouse amino acid sequence, the data cannot reveal anything about the relationships between, for example, bacteria and archaea, or plants and fungi.

As a whole group, ask your students the following questions:

- **What do you think is the significance of the random sequences?**
  The random sequences came from a random sequence generator. They represent how much amino acid identity you'd expect to find by chance (1/20 for each amino acid, or an average of 5%). Notice that although the % identity is fairly low in some organisms, they are still higher than

the 5% that would be expected by chance.

- **What *patterns* or trends do you see in the data?**
  Examples:
  Other animals tend to show the highest percent identity to mouse, followed by plants.

  The random sequences are the least similar to mouse.

- **What are some anomalies to the trends?**
  Example: For PCNA, the plant proteins are more like the mouse protein than the nematode (an animal) protein is.

- **What is something that "stands out," interests you, or raises further questions?**
  Examples:
  For topoisomerase, the proteins from archaea, plants, and bacteria all have a low percent identity to the mouse protein. Also, there is a big decrease between yeast and archaea.

  Cows, chimps, and chickens make Topoisomerase 1 proteins that are identical to mouse (at least for this portion of the protein).

- **Is there anything that you find "surprising"?**
  Example: RNA Polymerase protein from fungi (yeast and mold) are more similar to the mouse protein than the corn protein is.

- **How do you explain some of the patterns that you see?**

- **Do you think the patterns suggest that these organisms are related?**

Notes

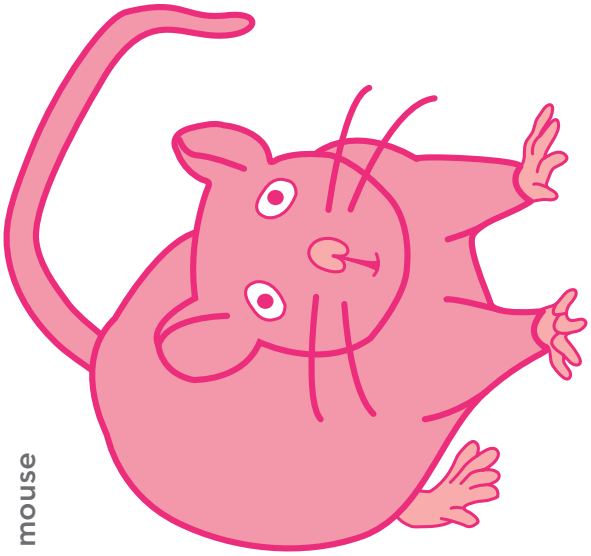
If students are having a hard time making connections or observing patterns, project the last 30 seconds or so of the Shared Functions, Shared Genes video (200-300 genes are common to all living things).

The organisms in this activity are also in the Online Phylogenetic Tree. Visit that activity to learn more about them.

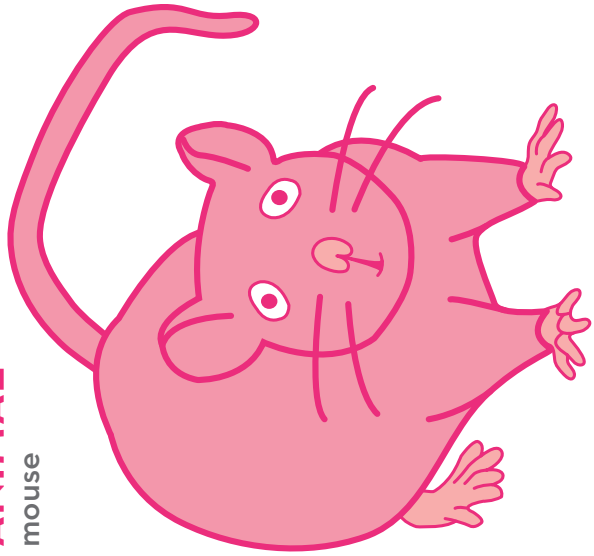
Random sequences were generated using Sequence Manipulation Suite: http://www.bioinformatics.org/sms2/random_protein.html

This material is based upon work supported by the National Science Foundation under Grant No. DRL-1418136. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

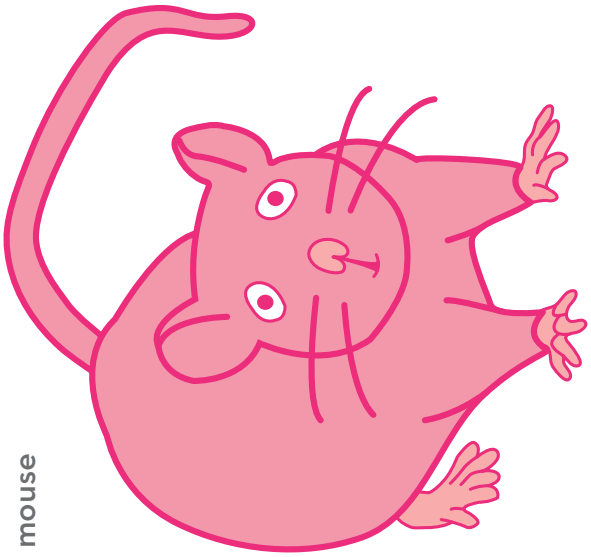
ANIMAL
mouse



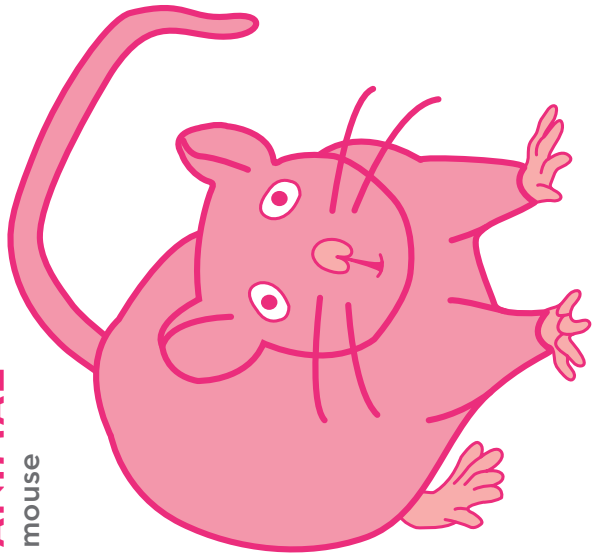
ANIMAL
mouse



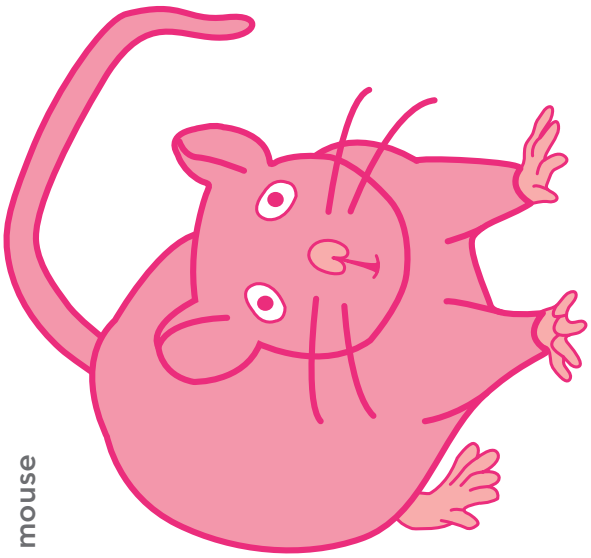
ANIMAL
mouse



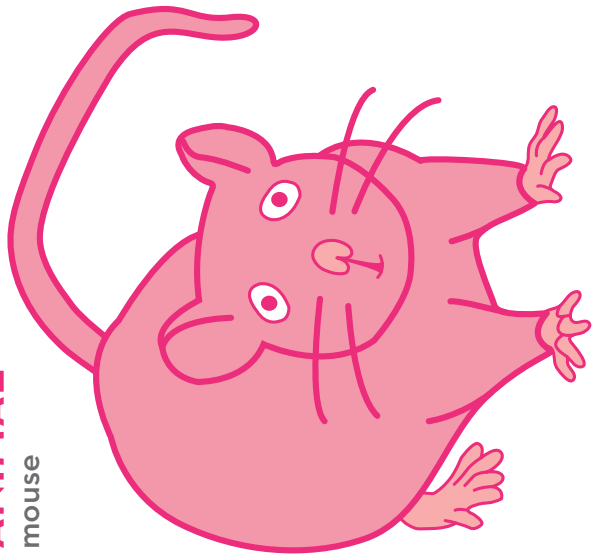
ANIMAL
mouse



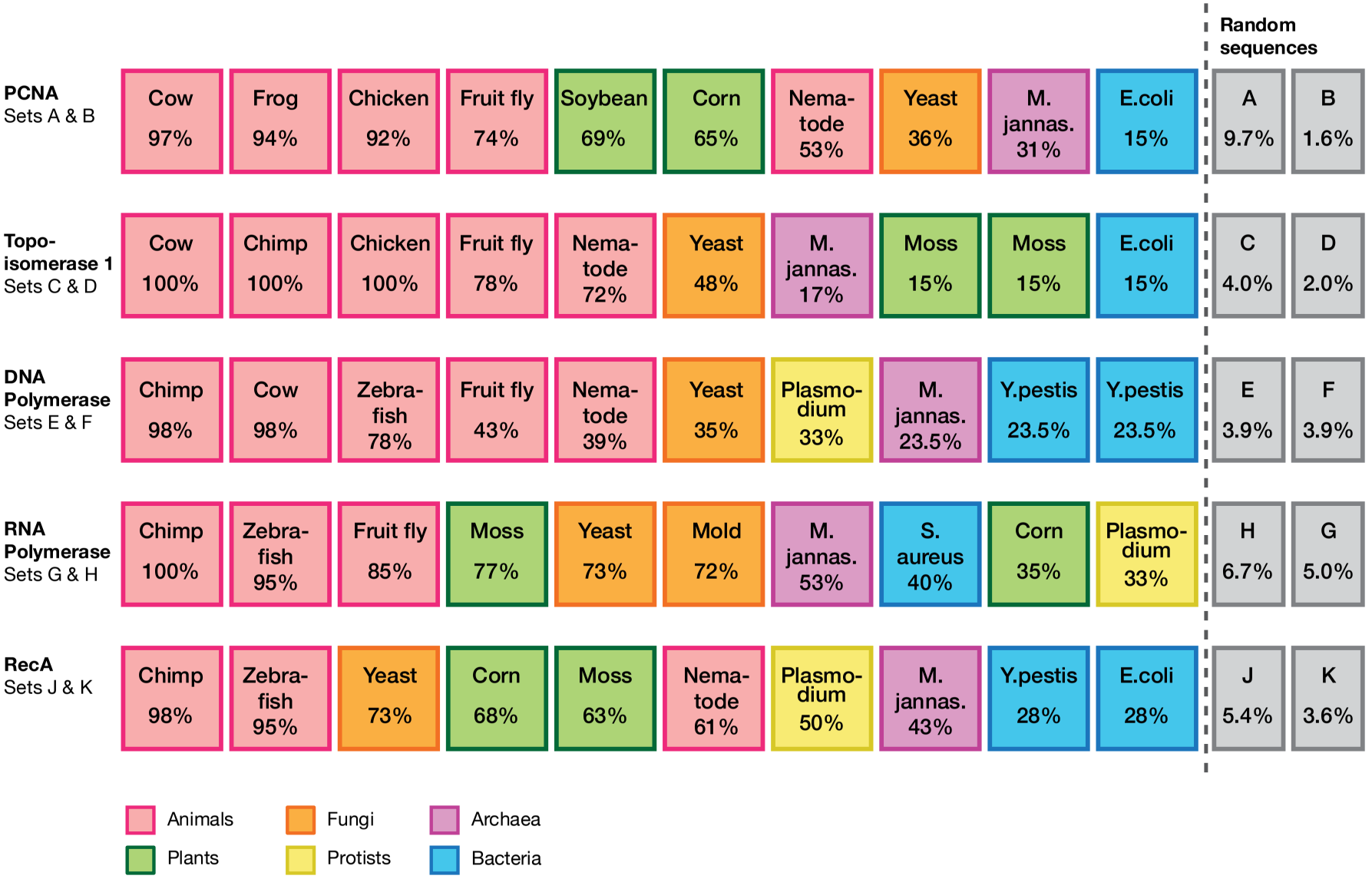
ANIMAL
mouse



ANIMAL
mouse



Key to Final Card Arrangement



Sequence references

Amino acid sequences were excerpted from the following NCBI reference numbers:

PCNA

mouse	NP_035175.1
chicken	NP_989501.1
fruit fly	NP_476905.1
soybean	NP_001241553.1
yeast	NP_009645.1
E. coli	WP_000673464.1
cow	NP_001029666.1
frog	NP_001081011.1
nematode	NP_500466.3
corn	NP_001105461.1
M. jannaschii	WP_064496904.1

DNA Polymerase 1

mouse	NP_032918.1
chimp	XP_016798596.1
zebrafish	NP_001292393.1
nematode	NP_001255109.1
plasmodium	XP_001351430.1
Y. pestis	WP_002210466.1
cow	NP_001192994.1
fruit fly	NP_536736.2
yeast	NP_014297.3
M. jannaschii	WP_064496646.1

RNA Polymerase 2

mouse	NP_722493.2
chimp	XP_016807157.1
fruit fly	NP_476706.1
yeast	NP_014794.3
corn	NP_043015.1
S. aureus	WP_000918667.1
zebrafish	NP_001019632.2
moss	XP_001766425.1
penicillium	XP_002568295.1
plasmodium	CAA36427.1 (GenBank)
M. jannaschii	WP_064496711.1

Recombination Protein A

mouse	NP_035364.1
zebrafish	NP_998371.2
nematode	NP_001023465.1
moss	CAC86603.1 (GenBank)
malaria	XP_001347762.2
Y. pestis	WP_002209446.1
chimp	XP_001144621.1
yeast	NP_011021.3
corn	NP_001104918.1
M. jannaschii	WP_064496644.1
E. coli	WP_000963153.1

Topoisomerase 1

mouse	NP_033434.2
chimp	XP_514649.2
nematode	NP_493337.1
moss	XP_001771420.1
E. coli	WP_001297122.1
chicken	NP_990441.1
fruit fly	NP_511161.2
yeast	NP_014637.1
M. jannaschii	WP_064496891.1