

Not Losing the Forest for the Trees: Learning to Compare Trees and Assess Support for Phylogenetic Hypotheses



School of
Education

Sam Donovan¹ and David Hornack²

1. Department of Instruction and Learning, School of Education

2. Department of Biological Sciences

University of Pittsburgh, PA 15260 sdonovan@pitt.edu



<http://www.bioquest.org/bedrock>

The Importance of Tree Reading

Too often, particularly in the context of evolution education, trees are treated as true phylogenies rather than hypotheses. Concomitant with this overemphasis on reifying a single tree, the emergence of molecular data is sometimes treated as a panacea for resolving phylogenetic disputes. An alternative approach to evolution education involves engaging students with some of the complexity involved in using multiple data sources to build trees and infer phylogenies. Building evolutionary trees from different data sets or using different tree-building methods will often produce different phylogenetic hypotheses about the relationships between taxa (Gura, 2000). Introductory biology instruction rarely addresses how to analytically read and compare trees as a way to look across multiple lines of evidence about evolutionary history. Working with multiple phylogenetic hypotheses is an important part of understanding the ways that evolutionary biologists use trees to reason about biological patterns and processes.

This poster presents a set of teaching resources built around the evolution of whales. The activities outlined here are designed to help students begin to address general issues of tree reading including comparisons of trees to identify areas of congruence and conflict, and understanding the differences between gene trees and species phylogenies.

Activity 1: Getting Started

The two trees below represent different hypotheses about the evolutionary relationships among the Cetacea (whales) and various ungulates (see Figure 1). Within the hoofed mammals there is a well defined split between those with an even number of toes called Artiodactyls which include pigs, hippopotamuses, llamas, cattle, deer and goats, and those with an odd number of toes called Perissodactyls which include horses, zebras, tapirs and rhinoceroses. Due to the extensive morphological adaptation that occurred in the whales' lineage it is difficult to place them within the ungulates by simply counting their toes. The evidence for these two evolutionary hypotheses (trees) involve comparisons of different sets of molecular and morphological data across the groups.

Q Describe several ways that these two hypotheses are similar and several ways they are different.

Q State in your own words the evolutionary relationships between Cetacea and Artiodactyls in each tree.

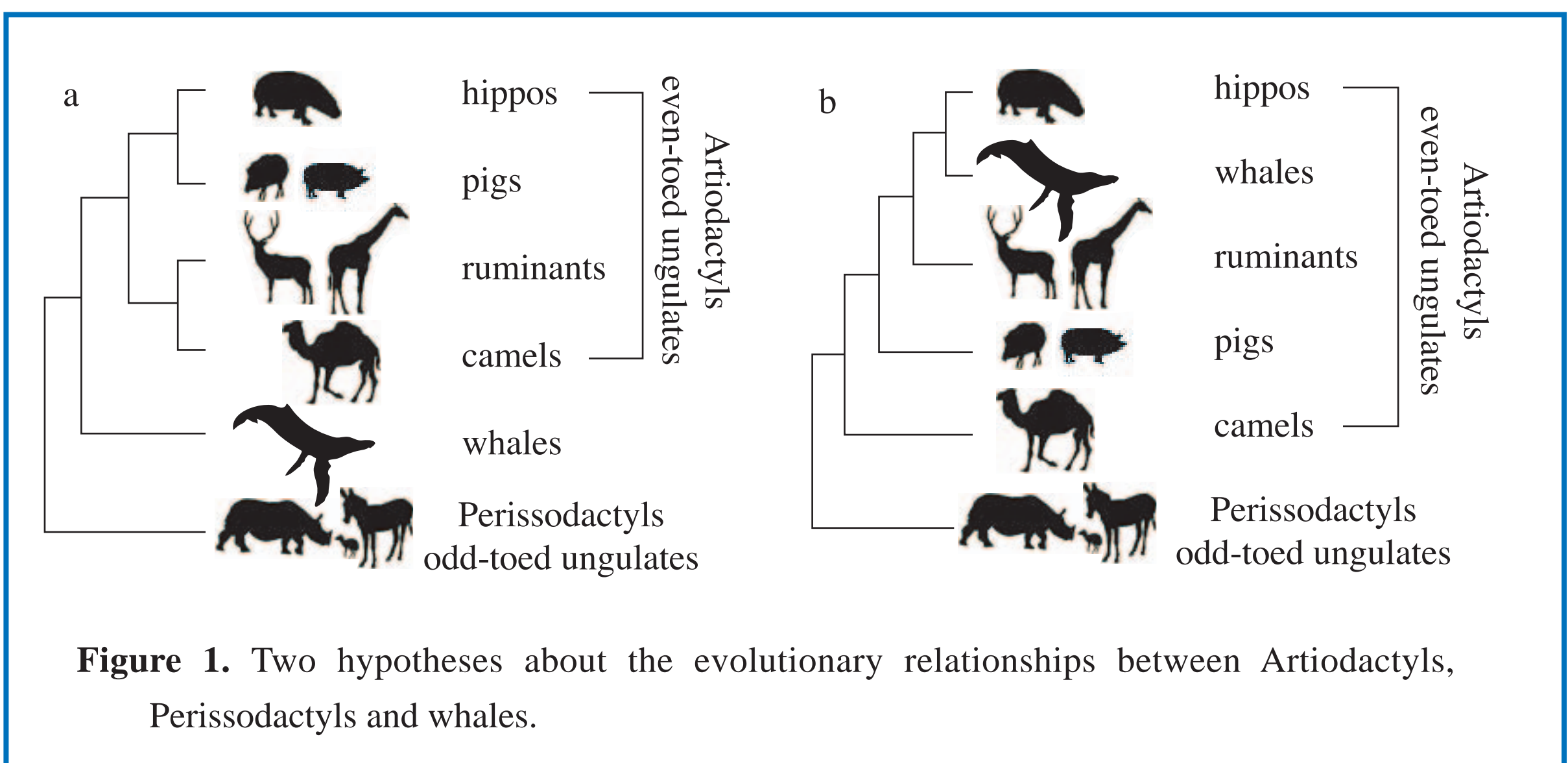


Figure 1. Two hypotheses about the evolutionary relationships between Artiodactyls, Perissodactyls and whales.

Supported by

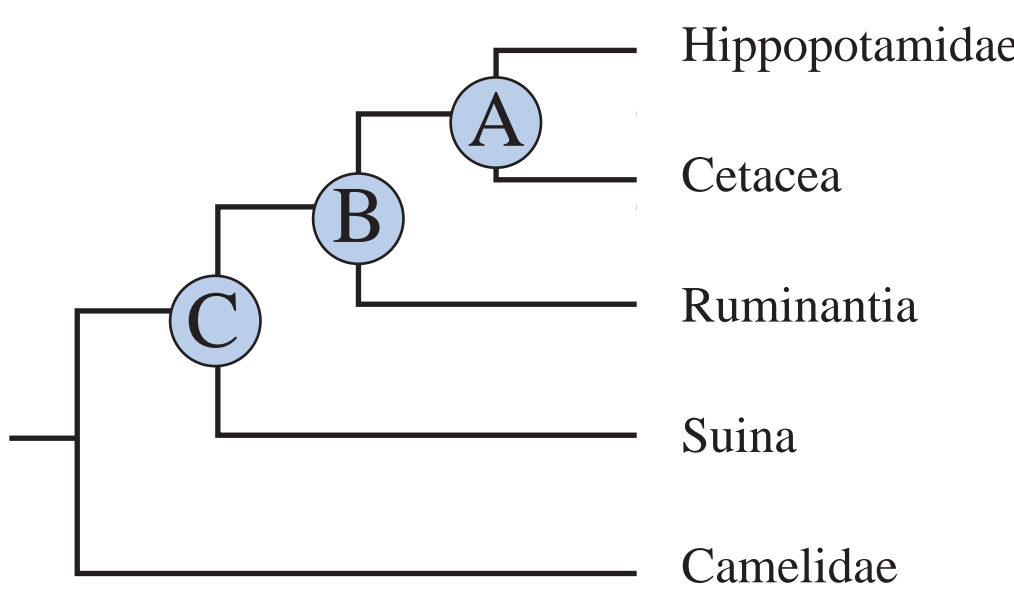


Activity 2: Hypothesis Testing

When looking at phylogenetic hypotheses about the evolutionary relationships among taxa biologists often must consider data from several sources. Part of making sense of multiple overlapping trees involves deciding where they are in agreement and where they are in conflict. For the questions below consider each of the four trees in Figure 3 as they relate to the hypothesized phylogeny presented in Figure 2.

Hypothesis

Figure 2. Tree based on an analysis of beta-casein gene data from Gatesy, et al. 1996.



Q Can you account for the different numbers of taxa present in each tree? How can you compare trees with different taxonomic information?

Q What is the hypothesis in Figure 2? What do the labels A, B and C represent? Is it possible to have support for clade B if clade A is not supported? How?

Q Which of the trees in Figure 3 support the hypothesis in Figure 2? Which support A? B? and C?

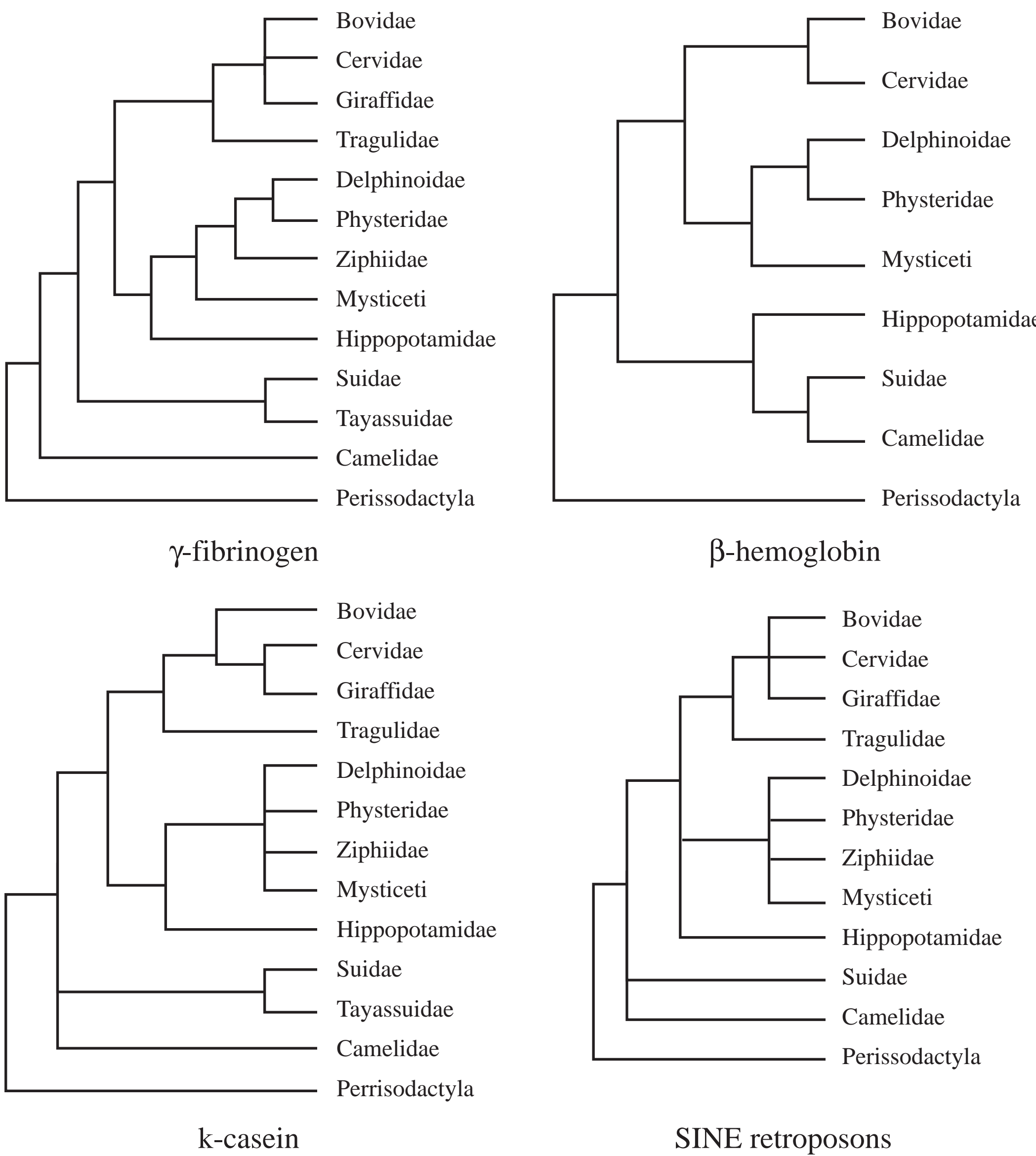


Figure 3. Four trees included in the Whippo-1 dataset and analyzed in Gatesy, et al. 1999.

Activity 3: Extended Investigation

Activities 1 and 2 are part of a larger collection of materials available in the Whippo Problem Space, a curricular resource that is part of the BEDROCK Bioinformatics Education Project. A problem space is a way of organizing diverse kinds of resources to support student inquiry. We have chosen to think about teaching and learning bioinformatics in the context of problem spaces to reflect some of the exciting possibilities and serious challenges that the flood of molecular data present for biology education. In contrast to a more traditional lab approach where the students may be asked to follow a highly structured series of procedures to confirm an experimental result, our view of biology education emphasizes the development and exploration of students' questions as they come to understand biological principles, analytical procedures, and the ways that inferences are made from the collection and analysis of data.

The Whippo Problem Space contains additional information about whale evolution, phylogenetic tree interpretation, and sequence data from the 15 molecular characters in the Whippo-1 data set (Gatesy, et al., 1999). The materials provided can be used to address general issues of tree reading, comparisons of trees to identify areas of congruence and conflict, details of using different types of molecular data, and as a launching point for more extended investigations of phylogenetic techniques.

Representatives from 13 taxa:

Bovidae (sheep, cattle, antelopes)	Camelidae (camels and llamas)
Cervidae (deer)	Delphinoidea (beluga whales, dolphins, porpoises)
Girafidae (giraffes)	Hippopotamidae (hippos)
Mysticeti (baleen whales)	Physteridae (sperm whales)
Suidae (pigs)	Tayassuidae (peccaries)
Tragulidae (chevrotains)	Zhiphiidae (beaked whales)
Outgroup (rhinos, horses and guinea pigs)	

17 data sets:

12S ribosomal DNA	16S ribosomal DNA	α -crystallin A
α -hemoglobin	α -lactalbumin	β -casein
β -hemoglobin	cytochrome b	cytochrome c
γ -fibrinogen	κ -casein	morphological characters
pancreatic ribonuclease	protamine P1	SINE retroposons
transposons 9	vWF (von Willebrand factor)	

Figure 4. Summary of the Datasets (Gatesy, et al., 1999).

Q How are trees built from distance analyses similar and different from the trees reported in the Gatesy paper?

Q Describe three different approaches to combining data from multiple genes to identify a species phylogeny. What biological processes could be used to explain the differences between the gene trees?

Q How do trees built from the amino acid sequences differ from those built with nucleic acid sequence? Do different genes show different rates of synonymous substitutions? How could that effect phylogenetic inference?

For more information visit:

<http://bioquest.org/bedrock/problem_spaces/whippo/>

Literature Referenced

- BioQUEST Curriculum Consortium. 2004 <http://bioquest.org/bedrock/problem_spaces/whippo/>
- Gatesy, J. et. al. (1999). Stability of cladistic relationships between Cetacea and higher-Level Artiodactyl taxa. *Systematic Biology* 48(1):6-20. Dataset available at: <http://hydrodictyon.eeb.uconn.edu/systbiol.org/issues/48_1/Gatesy1>
- Gatesy, J., Hayashi, C., Cornin, M. and Arctander, P. (1996). Evidence from milk casein genes that Cetaceans are close relatives of Hippopotamid Artiodactyls. *Molecular Biology Evolution*. 13(7):954-963.
- Gura, Trisha (2000). Bones, molecules...or both? *Nature* 406:230-233.