

# Investigating Evolutionary Questions Using Online Molecular Databases

Adapted from Puterbaugh and Burleigh, and *The American Biology Teacher* \*

## TEACHER BACKGROUND AND OVERVIEW

[Teacher Information Packet]

Although the morphological diversity of animals can generate student excitement, it is often harder to interest students in the use of molecular diversity to examine evolutionary relationships (molecular phylogenetics and systematics). The molecular revolution and the advent of computers give us greater power to assess relationships between organisms than just examining morphology. For this reason, we are in an exciting period of change in our understanding of relationships among organisms, and accordingly, it is worth the effort to persuade students that molecular systematics is an interesting field of study. Most scientific journals require that authors submit the molecular sequences reported in their scientific papers to public databases such as Genbank. The databases are therefore a rich source of data from thousands of studies performed throughout the world. These databases also have a great potential as a teaching tool. This series of simple exercises has three objectives:

- (1) to introduce students to the public molecular databases;
- (2) to engage students in addressing interesting morphological evolutionary questions with molecular data, and in the process introduce to students new concepts in both the molecular and morphological arenas; and
- (3) to illustrate important evolutionary concepts in a new and enjoyable way.

In the **first exercise**, students attack a report by Calvin (of the Calvin and Hobbes cartoon series, by cartoonist Bill Watterson) on “Bats are Birds” by comparing the amino-acid sequences of hemoglobin molecules from bats, birds, and non-bat mammal species (see the Editorial Comment #2 on page 3 of this packet). In the **second exercise**, students consider the evolution of whales from terrestrial animals. Scientists believe that whales evolved from ungulates (hooved mammals) that subsequently became entirely aquatic (Novacek, 1992). Students investigate whether the most recent land-dwelling hooved ancestor of the whales was more likely an artiodactyl (“even-toed” like sheep, hippos, cows, deer, and pigs), or a perissodactyl (“odd-toed” like horses, tapirs, zebras, and rhinos) ungulate. **Lastly**, the students use BLAST (Best Local Alignment Search Tool) to help them understand why some molecular systematists argue that the vertebrate class Reptilia should include birds. The term “paraphyletic” can be difficult to grasp, and in our opinion, this exercise illustrates the paraphyletic nature of the class Reptilia more clearly than any explanation. In other words, the exercise illustrates that Reptilia does *not* include all relatives of the most recent common ancestor of the group.

The exercises require some computer savvy to download the sequences from the Internet, but the evolutionary questions are fun and easy for advanced high school students or introductory college biology students to grasp. The exercises do not involve constructing phylogenetic trees since we wanted students to grasp the excitement of the questions without getting mired in the details of the methods.

[For Part II, if you would like your students to explore the discoveries and analyses of the evolution of whale morphology, see the “**Becoming Whales**” lesson on the ENSI site.]

[For Part III, a very useful article (“**Dinosaurs and Birds, an Update**”) by UC Berkeley paleontologist, specializing in bird evolution, would be worth reading here. It is available in its entirety on the ENSI web site.]

## Collaborative Science: Public Databases

Most scientific journals require that authors submit any sequences they use in a paper to a public, online database. Two frequently used online sequence databases are Genbank and Swiss-Prot. Genbank contains nucleotide data and is associated with the National Center for Biotechnology Information (NCBI), a part of the National Institutes of Health (NIH) (Benson et. al, 1999). Swiss-Prot contains only protein amino-acid sequences and is sponsored by the European Molecular Biology Laboratory (EMBL) and the Swiss Institute of Bioinformatics (SIB) (Bairoch and Apweiler, 2000). You can search these databases with descriptive words, scientific or common names, or with the special sequence identification numbers. In the exercises described here, students will search for amino-acid sequences of the hemoglobin beta chain from the Swiss database Swiss-Prot.

---

\* Adapted, with the kind permission of M. Puterbaugh and NABT, from Puterbaugh, M.N. & Burleigh, J.G. (2001). “Investigating Evolutionary Questions Using Online Molecular Databases.” *The American Biology Teacher*, **63** (6): 422-431, August.

## Using the Comparative Method to Answer Evolutionary Questions

The phrase “comparative method” describes a study that involves comparisons rather than experiments. When a researcher uses the comparative method, he/she neither manipulates nor controls conditions. The term often describes studies in which authors attempt to answer evolutionary questions by comparing numerous species in a given taxon. With few exceptions, any study that asks how traits arose among species must use the comparative method. Why? Speciation generally can not be observed in a human lifetime, and so experiments to manipulate conditions affecting speciation are difficult or impossible to conduct. The student exercises provided here are an introduction to the comparative method.

## The Hemoglobin Genes

Before one begins to make molecular comparisons between species, one must first choose a molecule to compare. The public databases such as GenBank and Swiss-Prot contain data on many different genes. In all the exercises, we use the beta chain of the hemoglobin molecule. We encourage teachers to develop their own exercises using other molecules as well. What makes the beta chain useful for teaching the comparative method? First, it is probably a molecule that students have encountered already. Second, the beta subunit is usually close to 146 amino acids long, and hence there are fewer problems aligning sequences from different animals than with proteins that vary more in length. Third, the public databases contain sequences from more than 100 different animals, including some bizarre and rare animals that may intrigue the students. For example, the Swiss-Prot database contains sequences from the Antarctic dragonfish, the electric eel, the yellow baboon, and the thick-tailed galago. If you do not know what the last animal is, try Part 4 looking it up elsewhere.

Students may already know that there is variation in the hemoglobin molecule among humans. Individuals who are carriers for sickle-cell anemia have one allele that makes a normal beta subunit and one allele that makes the sickle-type subunit. The sickle subunit differs by only one amino acid from the normal subunit. You may wish to begin these exercises by having students tell you what the percent similarity is between a normal and a sickle beta subunit. Because 145 amino acids are the same between the sickle-type beta hemoglobin and the normal-type beta hemoglobin, and one amino acid is different, the percent similarity is  $(145/146) \times 100$  or 99.3%.

## Limitations of the Comparative Method

In the process of completing these exercises, students are likely to see that using the comparative method to evaluate evolutionary questions is no simple task. When the molecular and computer revolutions were in their infancy, many scientists hoped that the entire tree of life, describing the relationships of all organisms, might eventually be reconstructed. The scientific community is now more aware of the limitations of molecular phylogenetics. Some of these limitations are listed below:

- 1) Lateral (or horizontal) gene transfer can confound our ability to use molecular data to predict relationships between organisms. Lateral gene transfer occurs when a gene is passed to an unrelated individual. A purely hypothetical example would be a case in which a retrovirus infected a coyote, moved from the coyote to infect a black bear, and in the process transferred coyote DNA to the black bear. In this hypothetical example, the DNA traveled laterally rather than vertically from parent to offspring. Lateral gene transfer is a serious problem for identifying relationships among bacteria that share genes regularly through conjugation, transduction, and transformation; however it is also potentially a problem among multicellular organisms.
- 2) Convergent evolution occurs even in molecules. Two molecules may be similar not because two species are related but because of selection. For example, the lysozyme proteins of leaf-eating langur monkeys and cows have some similarities. These similarities arose not because cows and langur monkeys share a recent common ancestor but rather because they both eat leaves, and selection has acted on both species in a similar way. The form of lysozymes in the langurs and cows enhances the digestion of plant matter and is characteristic of animals that digest their food in a foregut (Stewart *et. al*, 1987). The comparative method lacks the power to assign cause in the same way experimental methods can. The comparative method can determine if certain predications are upheld, but it can not determine why two species have similar sequences.
- 3) If the same gene or protein is different lengths in different species, difficulties can occur in trying to align those sequences. You may make different conclusions about the similarity of the sequences, depending on how you align the two sequences.

- 4) Comparisons can be hampered if molecules are either too variable or not variable enough. Different molecules are needed for comparing animals that diverged long ago than for those that diverged more recently. You might consider this particular limitation in Part II because a more variable molecule than the hemoglobin beta chain would be better for determining relationships of whales with perissodactyls or artiodactyls.

**Materials:** For each student or pair of students:

- 1) Lesson Background and Overview package
- 2) Packet of Figures (or make your own sharper master set by printing those screens directly, then cut and paste them together, configured as shown in the examples on this site)
- 3) Packet of Tables for Data Collection
- 4) Copies of two articles on Whale Studies
- 5) For each student or pair of students, one computer with access to the Internet and with a word-processing program and an Internet browser running simultaneously.
- 6) Specimens or pictures showing morphological features of bats, birds, whales, fish, crocodiles, and perissodactyl and artiodactyl mammals. (Hint: a field trip to the zoo with a list of animals that you will eventually download hemoglobin sequences from would be a great way to prepare for this exercise!)

### Summary

The teaching value of public online databases depends greatly on the quality of the questions that students use the databases to answer. Our experience is that the questions chosen for these exercises generate student interest. Even though the exercises are online and make use of molecular data, students walk away with a greater understanding of organismal biology, better appreciating, for example, the differences between artiodactyl and perissodactyl and the concept of convergent morphological evolution.

### References

- Bairoch A. & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, **28**, 45-48.
- Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., Rapp, B.A., & Wheeler, D.L. (1999). GenBank. *Nucleic Acids Research*, **27**, 12-17.
- Gingerich, P.D., Smith, B. H., & Simons, E.L. (1990). Hind limbs of Eocene *Basilosaurus*: evidence of feet in whales. *Science*, **249**, 154-157.
- Novacek, M. J. (1992). Mammalian phylogeny: shaking the tree. *Nature*, **356**, 121-125.
- Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., & Okada, N. (1997). Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature*, **388**, 666-670.
- Stewart, C.-B., Schilling, J.W. & Wilson, A.C. (1987). Adaptive evolution in the stomach lysozymes of foregut fermenters. *Science*, **330**, 401-404.

**Authors:** Mary N. Puterbaugh  
University of Pittsburgh at Bradford &  
The Allegheny Institute of Natural History  
Division of Natural Sciences  
300 Campus Dr., Bradford, PA 16701  
mnp1+@imap.pitt.edu

J. Gordon Burleigh  
Division of Biological Sciences  
University of Missouri at Columbia  
110 Tucker Hall  
Columbia, MO 65211

### Editorial Comments

1. A number of **minor changes** have been made to the original article by Puterbaugh and Burleigh, with the kind permission and assistance of Dr. Puterbaugh. These changes were made primarily to facilitate its application as a downloadable lesson, ready to duplicate and use in the classroom, providing a separate packet of tables and spaces for recording data, answering questions, and teacher checking. In addition, there have been some changes in online access addresses (URLs), and these (along with some alternatives) have been included here.

2. You may wish to prepare an overhead transparency of the **Calvin and Hobbes cartoon strips** mentioned in the lesson. You can find them at <<http://www.ultranet.com/~dtweed/calvin/w0726712.htm>>. They were published in 1989 (3 November and 30 October). You would need at least the 3 Nov. one, but the 30 Oct. strip provides a nice intro to the later one.

3. The **Figures** referenced in the student handout (Figures A-H, on four pages) are not nearly as sharp as would be desired. They do give one a good idea of what the various screen images look like, but are really too fuzzy as is. If you like, you could simply follow the sequence presented in the student handout, and get a printout of each key page mentioned as a figure. Then, prepare old-fashioned paste-ups, ideally with two figures on a page, make an 8.5 x 11 master copy from each page so prepared, and use those master to run off classroom sets of the figures.

4. Once your students have completed this 3 (or 4) part lesson, you might want to encourage them to explore **other biological questions**. For example:

- Are Giant Pandas more closely related to Lesser Pandas, or to bears? (students might try some different molecules to compare).
- Are the molecules of animals classified in the same taxon level (e.g. the same class, or same order, or same family) more alike than they are when compared to animals in the next higher level?
- Echinoderms (e.g., sea stars, sea urchins, and sea cucumbers) are considered more closely related to vertebrates than to other major invertebrate groups (e.g. molluscs and arthropods). Can you find a molecule which can be compared between individuals in these different groups, and does that comparison confirm the relationship based on morphological features? Consider comparing **lungfish**-coelacanth-other fishes with amphibians, or compare **peripatus** with (other?) arthropods and annelids, or **brachiopods** with some deuterostome and proterostome, to see which groups they are closest to. (May need to look at their cytochrome c or other common proteins to find something sequenced for all species to be used in the comparison).

5. This lesson provides a very nice extension **from** a lesson on whale evolution “**Becoming Whales**”, (perhaps more appropriate for high school / middle school levels) in which students “experience” the sequence of discoveries in which a number of transitional fossils gradually filled in much of the “gap” between the purported terrestrial whale antecedents and modern type whales, based on morphology. Other types of evidence are also explored, as are very recent comparisons of “whale ankles and heels” and their possible relationship to hooved animals (ungulates: perissodactyls vs. artiodactyls). <<http://www.indiana.edu/~ensiweb/lessons/whale.ev.html>>.

6. PHASE 2: An excellent extension of this entire experience is for your students to experience how scientists use comparison data (e.g. morphology, amino acid sequences, etc.) to build **evolutionary trees** as phylogenetic hypotheses, and the kind of parsimonious decision making which is involved here. It just so happens that such a walk-through is provided in the September 2001 *American Biology Teacher* (vol. 63, no. 7, pages 518-523): “The Comparative Method, Hypothesis Testing & Phylogenetic Analysis” by Fred Singer, Joel Hagen, and Robert Sheehy. Web sites where online data bases can be accessed are included.

7. We welcome, in fact encourage, your **feedback** from your experiences using these lessons. Your comments, questions and suggestions will be shared with Dr. Puterbaugh, and may well be reflected in future revisions of this lesson. Dr. Puterbaugh has indicated she appreciates any such feedback directly (via her e-mail or snail-mail addresses) if you prefer.

**Some additional references you could use as alternatives to those two listed in the student packet:**

- Adachi, J. *et al* (1993) Tempo and Mode of Mitochondrial DNA Evolution in Vertebrates at the Amino Acid Sequence Level: Rapid Evolution in Warm-Blooded Vertebrates. *J Mol Evol* 36:270-281.
- Gatesy, J. (1997) More DNA Support for a Cetacea/Hippopotaminidae Clade: The Blood-Clotting Protein Gene Gamma-Fibrinogen. *Mol. Biol. Evol.* 14(5):537-543.
- Graur, D. & D.G. Higgins (1994) Molecular Evidence for the Inclusion of Cetaceans within the Order Artiodactyla. *Mol. Biol. Evol.* 11(3):357-364.
- Hashimoto, T. *et al* (1993) The Giant Panda is Closer to a Bear, Judged by Alpha- and Beta Hemoglobin Sequences. *J Mol Ev.* (Contact author for vol. and pages; shows example of the use of hemoglobin in a published study.)
- Montgelard, C. *et al* (1997) Phylogenetic Relationships of Artiodactyls and Cetaceans as Deduced from the Comparison of Cytochrome b and 12S rRNA Mitochondrial Sequences. *Mol. Biol. Evol.* 14(5):550-559.
- Smith, M. *et al* (1996) Phylogenetic Evidence from the IRBP Gene for the Paraphyly of Toothed Whales, with Mixed Support for Cetacea as a Suborder of Artiodactyla. *Mol. Biol. Evol.* 13(7):918-922.
- Ursing, B.M. & U. Arnason (1998) Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. *Proc. R. Soc. Lond. B*, 2252-2255.