

MOLECULAR SEQUENCES & PRIMATE EVOLUTION: AN AMINO ACID EXAMPLE

Martin Nickels, Craig Nelson, Doug Karpa-Wilson, Steve Freedberg and Nathan Murphy

Student Procedures

INTRODUCTION

For many years, living organisms have been organized (classified) into nested groups based mainly on their anatomical similarities and differences. It is clear that members of “lower” groups (species, genera) are very similar in many respects, while members of “higher” groups (classes, phyla) are very different.

In addition, careful studies of a growing number of fossils have clearly shown us that life has greatly changed over the hundreds of millions of years that life has existed on our planet. Virtually none of today’s species existed 100 million years ago, and most of those that existed then are no longer with us. On the other hand, there are many fossil examples showing gradual changes over time from earlier forms to modern forms, and more are being found every year.

All of this indicates that common ancestors connect all life forms to each other, much like the way you are related to your cousins because you have the same grandparents. When this flow of life is traced backwards into deep time, we see many examples of gradual changes from earlier forms. This “descent with modification” is called *evolution*.

Recent studies using molecular biology (including studies of DNA, proteins, and biochemical processes) have revealed even sharper patterns of ancestry, mostly confirming the earlier patterns based on anatomy, but also revealing some connections that were not clear before. Furthermore, molecular methods are more easily quantifiable, so we can actually measure degrees of similarity (and therefore degrees of relationship), and we can do this with a wide array of biological molecules. As it turns out, these many independent efforts have widely confirmed and clarified the evolutionary connections postulated earlier.

This lesson provides an opportunity to peek into this molecular world, and see how it reveals ancestry. As an example, we will be comparing the amino acid sequences of beta hemoglobin molecules in eight different species.

DIRECTIONS: For each Section below, take turns in your team following the directions, discuss the questions, and answer them on a separate sheet of paper. Each person must be prepared to respond to these questions during classroom teacher-led discussion to follow.

SECTION A – Building a Matrix From a Set of Sequences.

This will help to process the raw sequence data into a format suitable for building a diagram showing relationships.

A hemoglobin molecule complex consists of two protein chains: an alpha chain and a beta chain. There are 146 amino acids in the beta chain of hemoglobin. On the **DATA SHEET**, you will find the amino acids listed for the beta hemoglobin sequences in eight different species. To save space, **ONLY** the 40 positions where the sequences are NOT identical across all eight species are shown; (the remaining 106 positions are identical for all 8 species). At the top of Worksheet A, you will find a partially completed matrix: “**Section “A” Matrix: Differences Among Amino Acid Sequences**”. You only need to complete the spaces in the upper portion of the matrix because the same numbers would go in the corresponding positions below the diagonal of “s’s” You will find an “x” in each redundant position below the diagonal. Each of the values in the matrix is the total number of positions where the amino acids differ between the two different species being compared.

1. What species comparison is represented by the first "s" in the matrix (i.e. species 1 is compared with which species)? How many differences would you expect for this comparison? Why? How many differences would you expect for all the other "self" (same species) comparisons along the diagonal? [s = self]

2. Two (or more) team members should count the differences between the sequences for species I and II, then I and III, then II and III. Enter the results in the three appropriate blanks in the matrix. Do your group members agree? If so, repeat for each remaining pair of sequences, comparing results, and completing the matrix.

3. Calculate and record the *average* values for each column (ignoring S and X).

SECTION B – Organizing Data [may be optional]

The rows and columns in the previous matrix were arranged to make the similarities apparent. When researchers first get actual molecular data, the data are not organized.

4. What would we have to do if (as in the matrix below) the data were *not* already organized for us? How would we do this? What would determine which two species we put in the first two columns? In the next column? [Compare the matrix below to the matrix produced in Part A].

	IV	VII	VI	I	II	V	III
IV	S	23	6	2	2	6	3
VII		S	22	25	25	22	24
VI			S	8	8	11	9
I				S	0	8	1
II					S	8	1
V						S	7
III							S

SECTION C – From Differences to Similarities

5. How many of the 146 amino acids in the beta chain of hemoglobin do the two most similar sequences **share**? How many do the two least similar sequences **share**?

6. Note that the values within each species' column are those for that species compared with each of the species to its left. What is the general pattern of differences among the column averages from left to right across the table? Are the values within each column usually more similar to each other than they are to the values in the other columns? What does this suggest about the biological relationships of each species to the species on its left?

SECTION D–Primate Similarities

SPECIES ID: The first seven species in the data table and matrix are *primates*

- | | |
|------------------------------------|---|
| I: Human | V: Rhesus Monkey (an Old World Monkey: OWM) |
| II: Chimpanzee (a Great Ape) | VI: Squirrel Monkey (a New World Monkey: NWM) |
| III: Gorilla (a Great Ape) | VII: Ring-tail lemur (a Prosimian) |
| IV: Common Gibbon (a "Lesser" Ape) | VIII: (to be revealed later) |

7. Place these names (or abbreviations) to the *left* of their Roman numerals in the Species column in Matrix A, and *above* their numerals that run across the top of the matrix.

8. Which group of *primates* is *least* similar to the others? (Prosimians, Old World Monkeys, New-World Monkeys, Lesser Apes, Great Apes or Humans?) Are the differences between this least similar group and the other groups all about the same, **OR** do they gradually increase from bottom to top, suggesting a “Ladder of Progress?” (The Ladder of Progress is a long outdated idea, in which species that branched off earlier haven’t changed. This would be indicated if, for example, the number of differences with gorillas were about half way between those for gibbons and those for humans.)

9. Are Gorillas more similar to Humans, or to Chimpanzees on the sequence data? Which species are most similar to Gibbons on the sequence data? Chimps and Gorillas are Great Apes. Should Humans be thought of as Great Apes as well? How could you explain these patterns in terms of common ancestry?

SECTION E – Building a Cladistic Tree (Cladogram)

The working assumption for the classification scheme known as *cladistics* is that every group of organisms arose by branching off from a previous group. Each branch is called a *clade*. That clade includes any and all subsequent branchings, so that one clade often includes many smaller clades. All the individuals in a clade share one or more carefully selected traits. Each trait must be identical or very similar within a clade, but appears to be modified (derived) from earlier (primitive, or ancestral) forms of the trait. Such *shared derived traits*, modified in different ways, distinguish the members of one clade from the members of another. The simplest diagram showing the sequence of dichotomous branches, based on the apparent sequence of modifications, produces a cladogram. Ideally, such a diagram follows the gradual accumulation of two or more traits (or their modifications) over time, showing a likely sequence of their evolution.

10. Which are the two species that have the most similar sequences (fewest differences)? Put these two species in the blanks indicating closest similarity (shortest branches, in the middle) on **Cladistic Tree A** (below the Section A Matrix). Does it matter which one you put on the left? Why?

11. What third species is most similar to the first two species? Put it in the appropriate place on Tree A. Select the next most similar species and put this fourth species appropriately on the tree. Why does this fourth species have nearly the same number of differences with each of the first three species? How would this relate to the time since the line leading to the first species diverged from the line leading to the fourth species?

12. Now put the remaining species on Tree A.

SECTION F – Analysis: Identifying Evolutionary Events from the DATA.

13. Look at position #33 on the DATA SHEET (not the tree). What amino acid did the common ancestor of old-world monkeys and apes probably have at this position? How can you tell from the sequence data along which branch (of Tree A) the amino acid sequence changed?

14. At what two positions does the Gibbon sequence (IV) differ from the African Great Ape sequences (II, III)? Given both the entire tree and the primate sequence data (I-VII), what mutation(s) happened at position 87 and along which branch(es)? At position 125?

15. What mutation(s) happened at position 50 and along which branch(es)? At position 104? (Note the ambiguity in each case!)

SECTION G – Cladistic Trees and Evolutionary Relationships.

Get WORKSHEET B with Cladistic Tree “B” and graph (at bottom) from your teacher.

Cladistic Tree A has been rearranged to form **Cladistic Tree B**. This was done by simply rotating each node (branching point) horizontally, starting at the lowest branching (1), then 2, 3, etc.

16. What is the difference in biological information (relationships, or branching sequence) between the two trees? (Ignore the dates in answering this question.) If there is no difference, explain why the trees look so different. Can you describe a way (e.g., a 3-D model or toy) to demonstrate this and make it easier to understand?

SECTION H – Introduction to Molecular Clocks.

The dates (in Millions of Years Ago = MYA) next to the nodes (branching points) of Cladistic Tree B represent the divergence (branching) dates based on fossil evidence and radiometric dating (not on molecular evidence).

17. Fill in the *average* number of molecular differences (as “**Changes**”) for each node on Tree B (using your numbers from the appropriate columns in the matrix table “A” (Part “A” Matrix). What is the general relationship between the geological ages of the nodes and the average numbers of molecular divergences? How could you explain this pattern? [For this reason, this pattern is called a “molecular clock.”]

18. On the Graph of **Amino Acid Sequence Differences vs Time** (below Cladistic Tree “B”), plot the data for the relationship between the *average* number of amino acid differences (changes) and the time since divergence (age of the node). What is the general relationship between the time and average number of differences? Why isn’t this relationship perfectly linear (i.e. why don’t the points all fall on a straight line)?

Species VIII in the data table and the matrix of differences is the *horse*. Horses, of course, are not primates and are not especially close to primates. The group that includes horses diverged from the primate lineage around 90 million years ago.

19. To Cladistic Tree B, add the branching line leading to the horse, and add the changes and time to the left of the branching node. Add the appropriate data point to the graph of **Differences vs Time**.

20. What does the horse data do to the shape of the graph curve? [Clue: A **back-mutation** is where a single DNA change produces an amino acid found in a more distant “cousin” at the same location in a molecule. How would back-mutations influence the shape of the graph curve? Why?]

21. Summarize what you learned with this lesson. Include what it suggests (or confirms) about human evolution and references to common ancestors. Also include how its results compared with observations of anatomy, chromosomes and skulls of hominoids (apes and humans).

Modified May 2005