
The Molecular Evidence for Evolution

Narrowing the Definition of Evolution to Represent a Single Concept

Robert Moss

The best way for students to discover how DNA sequences confirm descent with modification is to let them analyze data from the public sequence databases themselves. But first they must understand the significance of the sequences, and of variations in the sequences. The author communicates exercises he uses to help students understand the molecular evidence for evolution.

An article in a previous issue of this journal (*JCST*, Dec 1998/Jan 1999, pp. 167-170) begins, "Although there are several topics that give students concern, evolution is the most troubling subject area covered in biology courses."

This is true not only for students, but of those of us who teach evolution as well. We all have had difficulty discussing evolution with our classes; it often frustrates us as scientists and teachers that we cannot get the other side to "see" evolution the way we do. There are many reasons why we have not been successful in teaching evolution, and a number of things we can do to improve the situation.

WHAT IS EVOLUTION?

One of the major problems with our understanding of evolution is how we define it. "Evolution" and "natural selection" are terms used to explain processes in business, behavior, product development, and much, much more. Even most biologists use the term inappropriately, adding to the confusion.

Linhart recently studied biology

textbooks and found little agreement on a definition of evolution. The solution suggested by Linhart and others is to use the term *evolution* as broadly as possible; I believe we must *narrow* its use to represent a single concept, or "observation."

"Evolution," in the context most biologists intend to use it, should be defined simply as "descent with modification." That's the way Darwin used it, and that is the correct way. It is basically "incontrovertible." That's also the way the National Academy of Sciences' report, "Teaching About Evolution" defines evolution (report can be found online at: <http://www.nap.edu/readingroom/books/evolution98/>).

Darwin did *not* comment on the origin of life in his published works, and we need to stop linking the concept of the biochemical origin of life with the term "evolution." The fact is we have very little real, tangible evidence supporting the prevailing hypothesis of how life originated. And even if we *could reproduce* a biochemical origin of life, proving it to be possible, we cannot go back and *see* if life actually originated in that way. In contrast, we have much evidence supporting the idea that all the life forms on this planet arose from a common ancestor.

THE EVIDENCE IS COMPLEX

Sinclair and Pendarvis' article reports on a survey that they took of biology students' understanding and beliefs concerning evolution. The authors report that when asked what was the strongest supporting evidence of evolution, "Most did

not select the more recently identified evidences such as molecular, biochemical, genetic or embryological similarities even though these topics were given equal coverage during the instruction . . . we feel that students did not choose the more complex evidences because they lacked the background to adequately comprehend them."

It is difficult to explain the power of DNA sequence comparisons to college students who have never sequenced DNA for instance, and we often do not take the time required to try to make it understandable. I give my students DNA or protein sequence data to analyze. But first, I have to give them the tools they need to compare the sequences and then to interpret their observations. To begin, I use the following analogy:

Imagine that I gave a class of 10 students an assignment to write 10,000 word essays on a certain topic. On the due date, I get back 10 identical essays. What could I conclude? Well, the chances of each of the 10 students coming up with the identical essay independently are . . . well, basically zero. It is virtually impossible, so I know the essays were not written independently, but I do not know how they were written. I do not know who wrote the original paper, and who copied from whom.

Robert Moss is an associate professor in the department of biology, Wofford College, 429 N. Church Street, Spartanburg, SC 29303; e-mail: MOSSRE@WOFFORD.EDU.

GENE EVOLUTION EXERCISE

Below you will find the amino acid sequence of the myoglobin gene from 17 species of animals. Each letter represents one of the 20 amino acids, - represents an amino acid that is missing due to a deletion. Species #1 is human, and the protein has 153 amino acids. I've bolded all letters that differ from the human sequence.

Remember that the longer since two species have diverged, the more time they have had to accumulate mutations. Also remember the "copied essay" analogy I gave you in the beginning of the course. Based upon these, draw an evolutionary tree for species 1-17

1	11	21	31	41	51	61	71
1 GLSDGEWQLV	LN V WGKVEAD	IPGHGQEVLI	RLF K GHPETL	EKFDKFKHLK	SEDEMKASED	LKKHGATVLT	ALGGILKKKGH
2 GLSDGEWQLV	LN V WGKVEAD	IPGHGQEVLI	RLF K GHPETL	EKFDKFKHLK	SEDEMKASED	LKKHGATVLT	ALGGILKKKGH
3 SLSAAEADLA	GKS WAPVFAN	KN ANG L DFLV	AL FEKFPDSA	NFF ADFKGKS	VAD I-KASPK	LRD VSSRIFT	RL NEFV N NAAN
4MGLSDGEWQLV	LN V WGKVEAD	V AGHGQEVLI	RLF K GHPETL	EKFDKFKHLK	SEDEMKASED	LKKHGNTVLT	ALGGILKKKGH
5 GLSDGEWQLV	LN V WGKVEAD	L AGHGQEVLI	RLF K GHPETL	EKFDKFKHLK	SEDEMKGSED	LKKHGNTVLT	ALGGILKKKGK
6 GLSDGEWQLV	LN I WGKVEAD	L VGHGQEVLI	RLF K GHPETL	EKFDKFKHLK	SEDEMKRSED	LKKHGKTVLT	ALGGILKKKGH
7 GLSEGEWQLV	LN V WGKVEAD	L AGHGQDVLI	RLF K GHPETL	EKFDKFKHLK	TEA EMKASED	LKKHGNTVLT	ALGGILKKKGH
8 GLSDGEWQLV	LN I WGKVETD	E GGHGKDVLI	RLF K GHPETL	EKFDKFKHLK	SEDEMKASED	LKKHGITVLT	ALGNILKKKGH
9 VLSDAEWQLV	LN I WAKVEAD	V AGHGQDILI	RLF K GHPETL	EKFDKFKHLK	TEA EMKASED	LKKHGNTVLT	ALGGILKKKGH
10 GLSDGEWELV	L KTWGKVEAD	IPGHGEFVLV	RLFTGHPETL	EKFDKFKHLK	TE GEMKASED	LKKQGVTVLT	ALGGILKKKGH
11 GLSDGEWQLV	L KVWGKVEGD	LPGHGQEVLI	RLF K THPETL	EKFDKFKGLK	TE DEMKASAD	LKKHGGTVLT	ALGNILKKKGQ
12 GLSDGEWQLV	LN V WGKVEAD	L AGHGQEVLI	GL FKTHPETL	DK FDKFKNLK	SE EDMKGSED	LKKHGCTVLT	ALGTILKKKGQ
13 GLNDQEWQQV	L TMWGKVESD	L AGHGHAVLM	RLF K SHPETM	DR FDKFRGLK	TP DEMRGSED	M KKHGVTVLT	-LGQILKKKGH
14 -ADFDA-V	L KCWGPVEAD	Y TTMGGLVLT	RLF K KEHPETQ	K LFPKFAGIA	Q ADIAGNAA-	I SAHGATVVK	K LGELLKAKGS
15 -HDAEL-V	L KCWGGVEAD	F EGTGGEVLT	RLF K QHPETQ	K LFPKFVGLA	S NELAGNAA-	V KAHGATVVK	K LGELLKARGD
16 -TEWEH-V	N KVWAVVEPD	IPAVGLAILL	RLF K EHKETK	D LFPKFKEIP	V QQL-GNNED	LRKHGVTVLR	ALGNILKQKQK
17 GLSDQEWQQV	LTIWGKVEAD	I AGHGHEVLM	RLF H DHPETL	DR FDKFKGLK	TP DQMKGSED	LKKHGATVLT	Q LQKILKQKGN
82	92	102	112	122	132	142	152
1 HEAEIKPLAQ	SHATKHKIPV	KYLEFISECI	IQVLQSKHPG	DFGADAQQAM	NKALELFRKD	MASNYKELGF	QG
2 HEAEIKPLAQ	SHATKHKIPV	KYLEFISECI	IQVLH S KHPG	DFGADAQQAM	NKALELFRKD	MASNYKELGF	QG
3 AG KMSAMLSQ	FA KEHVGFGV	G SAQFENVRS	M FPGFVASVA	A PPAGADA A W	T KLFLI I D	A LKAAGK	
4 HEAELTPLAQ	SHATKHKIPV	KYLEFISEAI	IQVLQSKHPG	DFGADAQQAM	S KALELFRND	M AAKYKELGF	QG
5 HEAELKPLAQ	SHATKHKIPI	KYLEFISEAI	IQVLQSKHPG	?FGADAQQAM	K RALELFRND	I AAKYKELGF	QG
6 HDAELKPLAQ	SHATKHKIPI	KYLEFISEAI	IHVLSKHPG	DFGAD T HAAM	K KALELFRND	I AAKYRELGF	QG
7 HDAELKPLAQ	SHATKHKIPI	KYLEFISEAI	IHVLSRHPA	EFGADAQQAM	NKALELFRKD	I ATKYKELGF	HG
8 HEAELKPLAQ	SHATKHKIPV	Q LEFISDAI	IQVIQSKHAG	N FGADAQAAM	K KALELFRHD	M AAKYKEFGF	QG
9 HEAELKPLAQ	SHATKHKIPI	KYLEFISDAI	IHVLSRHPA	DFGADAQAAM	NKALELFRKD	I AAKYKELGF	QG
10 HEAEIQPLAQ	SHATKHKIPI	KYLEFISDAI	IHVLSKHPA	EFGADAQQAM	K KALELFRND	I AAKYKELGF	QG
11 HEAELKPLAQ	SHATKHKISI	KFLEYISEAI	IHVLSK H SA	DFGADAQAAM	G KALELFRND	M AAKYKEFGF	QG
12 HAAEIQLAQ	SHATKHKIPV	KYLEFISEII	IEVKKRHS G	DFGADAQQAM	S KALELFRND	I AAKYKELGF	QG
13 HEAELKPLSQ	THATKHKVPV	KYLEFISEAI	M KVIAQKHAS	N FGADAQEAM	K KALELFRND	M ASKYKEFGF	QG
14 HAAILKPLAN	SHATKHKIPI	N NFKLISEVL	V KVMHEKAGL	DAGG-QTAL	R NVMGIIAD	L EANYKELGF	SG
15 HAAILKPLAT	THANTHKIAL	N NFRLITEVL	V KVMAEKAGL	DAGG-QSAL	R RVMDVVIGD	I D T YYKEIGF	AG
16 HSTNVKELAD	THINKHKIPP	K NFVLITNIA	V KVLTEMYP S	D MTGPMQESF	S KVFTVICSD	L ETLYKEANF	QG
17 HESELKPLAQ	THATKHKIPV	KYLEFISEVI	I KVIAEKHAA	DFGADSQAAM	K KALELFRND	M ASKYKEFGF	QG

Now imagine that I get back 10 papers that are almost identical (see figure 1 on page 131). I get one superb paper from John; James' paper is identical with one exception: word number 300 is different, and the sentence containing the difference now makes no sense at all. It seems quite likely that James copied from John [or from someone else who copied

from John], but made a mistake in word #300.

Jan's paper is identical to James', except it has two nonsensical mistakes . . . the very same mistake at word #300, plus another at word #1,100. And Susan's paper shares the very same mistakes at words #300 and #1,100, plus a new one at word #124.

How would you interpret this? Well first of all, the papers were not written independently. Even with the few differences, they are far too similar. Could they all have been copied from the same source, or all copied directly from John? Unlikely. The chance of Susan independently making exactly the same mistakes, at words #300 and 1100 that Jan made,

KEY:		
Species	Identity	# Amino Acid Differences from Humans
1	Human	
2	Mammal: primate chimpanzee	1
3	Molusc: slug sea hare	123
4	Mammal: pig	11
5	Mammal: otter	15
6	Mammal: sea lion	19
7	Mammal: porpoise	20
8	Mammal: kangaroo	22
9	Mammal: whale	23
10	Mammal: elephant	23
11	Mammal: platypus	24
12	Mammal: mouse	26
13	Bird: penguin	43
14	Fish: tuna	84
15	Fish: carp	88
16	Fish: shark	90
17	Bird: chicken	36

ANALYSIS OF THE DATA

The students are expected to count up the number of differences each species has when compared with the human sequence, and draw a simple tree, where the divergence points reflect the number of differences in the amino acid sequences. After returning the graded assignments, we break the code, and examine their trees. We find that the chimpanzee has only 1 difference when compared to human. The other mammals have 11-26 differences; birds differ by 36-43, fish by 84-90, and the mollusc by 123. The number of amino acid differences is proportional to the time since divergence of the species. These results fit well with the fossil record, and other types of evidence. A sample tree is shown in **Figure 2**.

A few other comparisons are easily made. Species 14, 15, and 16 share two deletions in the first 10 amino acids. It should be immediately clear that these three species must be closely related, and must share a common ancestor that had already diverged from the other species. The alternative would be that the deletions would have occurred independently, by chance, at the same place in the different species. And since species 14 and 15 share other deletions at amino acids 60 and 126, it is reasonably clear that 14 and 15 are more closely related than either is to 16. The maps can be adjusted accordingly, by counting the differences between closely related species. A tree including a few such comparisons is shown in **Figure 3**.

Many authors, including the Sinclair and Pendarvis in the article cited, suggest that the best way to approach evolution is by using "discovery," letting students investigate the evidence themselves. This approach is certainly quite powerful in helping students to discover and understand the genetic basis of descent with modification. Indeed, after the introduction of these exercises, my students no longer openly disagree with this point...I can only hope this means that the data has eliminated many of their uncertainties. I am now able to discuss evolution throughout my genetics course, without any of the open disapproval and disbelief I previously had to wade through. Students discuss evolution with me much more confidently as well. I'm convinced that these exercises have helped the majority of my students accept at least the basis of evolution, genetic descent with modification.

is extremely small. And extending the analysis, the chance of the tenth student independently making exactly the same eight errors as the previous student is basically zero. The only reasonable explanation is that James copied from John, then Jan copied from James, etc.

Well, the same thing is true when we compare genetic sequences of different organisms. Yes, their DNA is surprisingly similar . . . but it is the *differences*, the mutations shared by species believed to be close on the evolutionary tree, but not by more distant ones, that are the most instructive. And we are not talking about just a few instances, but thousands that have been examined.

The only reasonable way to interpret these differences is to assume that *all the species examined came from one another, "copied" in series, in a definite order* that we can now elucidate. Now, notice this evidence says nothing about *how* they were copied, whether by divine intervention or natural selection. But it shows quite definitively that all the species (papers) were created along a time line, one from another. And although this essay exercise is a rather simplistic analogy to the genetic sequence evidence, it does indeed help to understand a serial "copying" of papers or species.

It is important to note that this analogy is clearly an oversimplification. It is not a very good analogy for evolution as a whole. We do not have selective pressures acting on the essays the way they would act on genetic variation. This exercise discusses only "nonsensical" mistakes. And although geneticists often look at areas of DNA that have no meaning, such as introns and intergenic regions, the analogy applies even when we discuss alterations in coding regions, or the more rare "beneficial" mutations that provide a selective advantage to an organism. The analogy is designed simply to help the students understand how the differences and similarities in genetic sequences validate descent with modification.

Before we dive into the data, I use

one more analogy to help the skeptics to at least look at the data with an open mind. We "recreate" evolution, at least in our minds. Let us imagine for the moment that there is life on another planet. Of course, right now we have no direct evidence that there is life on any other planet. But let's pretend.

Let us imagine the simplest form of life, such as a single-celled organism not very different from our own bacteria perhaps, is living on well, let's say Mars. Would that life "evolve"? To answer that question using Darwinian evolutionary theory, we need to ask two others . . . would this life form demonstrate "variation," and would it experience "natural selection"?

The first is rather easy. Either there is variation [mutation], or the cells copy themselves perfectly, every single generation, with absolutely no mistakes. Given entropy, this would be impossible. Mistakes are inevitable. True, mistakes might be far less frequent than on earth, or far more frequent . . . but they must occur.

Given variation, would there be "natural selection" for the fittest organisms? Let us again look at the alternative. The only way the fittest will not outperform the less fit is if there's no competition. In other words, there's an unlimited supply of food, space, and resources, to accommodate an unlimited number of offspring. Again, that is not possible. As the population grows, there must come a point when the resources become limiting. Thus, there must be competition, and if there's competition, clearly some variants must survive more often or longer than others. That's natural selection.

When you put those two together, the inevitability of variation, and of eventual competition, you see that evolution must occur, or all life on the planet must die out. This argument says nothing about the rate of evolution, nor its results. It would be extremely unlikely for evolution on another planet to follow any "directions" similar to those of earth, since variation is due to chance mutations, and selective pressures in different environ-

continued on page 131

The Molecular Evidence of Evolution continued from page 113

ments are bound to be different as well. There would be no reason to expect humans, or bipeds, or even multicellular (or even cellular for that matter) organisms to evolve. But change, to forms better adapted to their environments, is inevitable.

THE MOLECULAR EVIDENCE

Hopefully, now the students are ready to dive into the data. We might analyze DNA sequences directly; however I find it more convenient to look at the amino acid sequences of proteins, which of course reflect the coding sequences of the DNA. The assignment as given to my students is shown on page 112; the protein sequences

come from the GENBANK web site: www.ncbi.nlm.nih.gov/Entrez/index.html [note the address is case sensitive].

Acknowledgment

The author acknowledges with thanks the assistance of Dr. Peter Cherbas of Indiana University, who sent him

15 of the 17 sequences shown, all originally from Genbank.

References

- Sinclair, A., and M. Pendarvis. 1997/98. Evolution vs. conservative religious beliefs. *Journal of College Science Teaching* 27(3): 167.
- Linhart, Yan. 1997. Education: The teaching of evolution—We need to do better. *Bioscience* 47(6): 385.

