J. Mol. Evol. 3, 179—188 (1974)

Evolutionary Clock: The Rate of Evolution of Rattlesnake Cytochrome *c**

David Penny

Department of Botany and Zoology, Massey University, Palmerston North

Received July 15, 1973

Summary. It is shown that the method used by Jukes and Holmquist [Science **177**, 530 (1972)] is not able to lead to any conclusion on the rate of evolution of rattlesnake cytochrome *c* because the method does not consider the time period over which the observed differences occurred. In an attempt to overcome this problem the phylogenetic relationship between rattlesnake, turtle and birds is examined from the paleontological evidence and from phylogenetic trees constructed from cytochrome *c* sequences by "matrix methods" and by "ancestral sequence" methods. The paleontological evidence and the "ancestral sequence tree" are in agreement for the positioning of rattlesnake. This ancestral sequence tree is used to estimate the rate of amino acid substitution and minimum base changes for different lines of descent among 20 vertebrate species. The rate of amino acid substitution is faster than average on the rattlesnake line but is not the fastest among the vertebrates and it is concluded that no "species specific" effect has yet been demonstrated for rattlesnake. However there is a large amount of diversity in the rates of amino acid substitution and this is discussed from the concept that at any point in time only a few codons (the covarions) are able to accept an amino acid substitution. It is suggested that some fluctuations in the rate of amino acid substitutions should occur.

Key words: Covarions — Cytochrome *c* — Evolutionary Clock — Molecular Evolution — Rattlesnake — Vertebrates.

It has recently been suggested by Jukes and Holmquist (1972) "that the evolutionary rate of change of cytochromes *c* is species dependent as well as time dependent" and that "the evolutionary clock does not run at the same rate for all species". The basic method used by the authors involved choosing the two published reptilean cytochrome *c* sequences (rattlesnake and snapping turtle) and the four different bird cytochrome *c* sequences. The differences in sequences is calculated between such groupings as birds-turtle; turtle-snake, and birds-snake.

Three objections to this method can be considered. The main objection is that it is assumed that the observed differences have occurred over the same time period but no evidence for this is presented. The second objection is that conclusions are dependent on the species chosen for the comparison

* This study was completed while the author was on leave at the Biology Department, Carleton University, Ottawa, Onatrio, Canada.

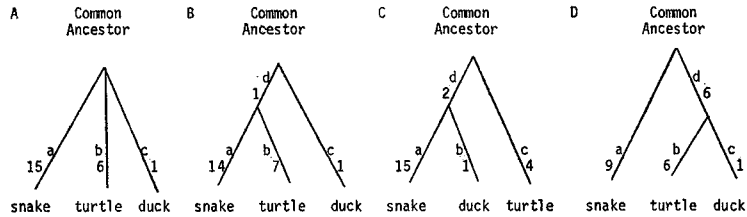
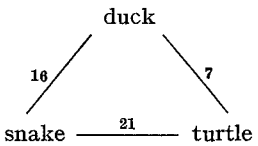


Fig. 1 A—D. Possible phylogenetic arrangements of rattlesnake, turtle and duck. One possible set of solutions is given for the number of changes to the amino acid sequence on each branch that would give 21 differences between snake and turtle; 16 between snake and duck and 7 between turtle and duck. In general parallel (repeated) or reverse mutations are ignored except in B where for example the mutation on *d* could have mutated back to the original on *b* to give 7 differences between turtle and duck. Either B, C or D would reduce to A if there were no changes on the branch *d*

and in particular justification was not given for excluding the mammals although at least two published phylogenetic trees derived from cytochrome *c* sequences (Fitch, 1971a; Strydom, van der Walt and Botes, 1972) show rattlesnake separating before the divergence of mammals from the birds and turtles. The third and perhaps a less fundamental objection is that the authors appear not to have explicitly considered the possibility that either birds or turtle may have evolved abnormally slowly rather than rattlesnake having evolved unusually fast.

It is clear that to establish a rate of amino acid substitution it is necessary to know both the number of changes to the sequence and the time period over wich the changes have occurred and Fig. 1 is chosen to illustrate this more clearly. To simplify the illustration only one bird sequence is used. The number of differences between the cytochrome *c* sequences of rattlesnake, turtle, and duck is (Dayhoff, 1972);



and the four possible phylogenetic trees are given in Fig. 1. On each line of descent is given an example of the number of changes that could give the observed differences.

If either A, B or C of Fig. 1 is the correct representation of the phylogeny then the conclusion is valid that rattlesnake has evolved faster than the other but it would still not be possible to decide whether turtle and/or duck had changed abnormally slowly. If, however, D is the correct representation then no conclusion about the rates of evolution is possible since all that can be said is that the lengths $a+d=15$. The figures shown for *a* and *d* in D would give a rate of evolution for snake cytochrome *c* between that of

turtle and duck but the correct rate could be either faster or slower than the other two. Other species can be substituted in Fig. 1 using the available data from Dayhoff (1972). If for example penguin, duck and turtle are compared the method would lead to the conclusion that turtle cytochrome *c* had evolved twice as fast as the other two. This example illustrates the second objection to the method that the results are not independent of the species chosen for comparison.

Phylogenetic Trees

At this point it is clear that the argument used is not valid but it is possible that the conclusion could be justified by some alternative method. The following discussion is an attempt to distinguish between the alternatives A—D and if necessary to have an estimate of the length of the branch *d*. Evidence can be obtained from phylogenetic trees constructed either from cytochrome *c* sequences or from the standard evidence of comparative anatomy and paleontology. The fossil record shows the bird line (and crocodiles and dinosaurs) and the snakes (via lizards, Rynocephalians (tuatara) and Eosuchians) to be separate lines since the late Permian (Romer, 1966) and the time of divergence is probably at least by the middle Permian about 250 million years ago (Harland, Gilbert, Smith and Wilcock, 1964; York and Farquhar, 1972). The turtles have a distinct fossil record to the Triassic but the origin of these forms is uncertain (Romer, 1966). It was thought possible that the turtles were linked to the Cytolosaurs through an incompletely known Permian form *Eunotosaurus* (Romer, 1966) but a more recent examination of this fossil by Cox (1969) concludes that *Eunotosaurus* is not ancestral to the turtles and this view is accepted by Romer (personal communication). It is therefore not possible at present to decide between alternatives A—D from the palentological evidence although it would favour alternative C.

The two main approaches that have been used to construct phylogenetic trees from amino acid sequence data are the "ancestral sequence" method (Dayhoff, 1972; Boulter *et al.*, 1972; Ramshaw *et al.*, 1972) and what has been described by Dayhoff (1972) as "matrix methods". The two methods give similar but not always identical results (Dayhoff, 1972) and each method will be considered. Two "matrix methods" phylogenetic trees that include rattlesnake have been published (Fitch, 1971a; Strydom *et al.*, 1972; Fitch and Margoliash, 1968) although one of these (Fitch and Margoliash, 1968) does build ancestral sequences for the tree with the lowest "percentage standard deviation". Both of these studies conclude that the best tree for reptiles, snakes and birds is Fig. 1D and indeed both have rattlesnake separating prior to the divergence of mammals from birds and reptiles. In this respect these trees disagree with the paleontological evidence.

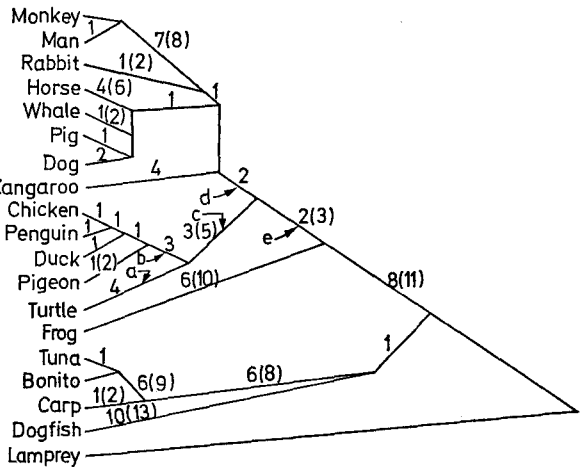


Fig. 2. A phylogenetic tree constructed by an ancestral sequence method. Nineteen vertebrate cytochrome *c* sequences are taken from Dayhoff (1972) except that by comparison with bonito positions 60 and 61 of tuna are reversed to Asp-Asn and the rattlesnake sequence as modified by Jukes and Holmquist (1972) is used. The justification for this can be seen in Dickerson (1972). If no change to a particular amino acid occurred more than once at a given site and if there were no reverse changes, then there would have to be 68 amino acid changes or 76 minimum base changes. The phylogenetic tree shown is one of several that requires 23 additional or "duplicated" amino acid changes and "duplicated" base changes. Rattlesnake is then added to each of the 37 $(2n-1)$ possible positions and the number of "duplicated" changes calculated for each tree. With this sequence for rattlesnake there would need to be a minimum of 79 amino acid changes or 87 minimum base changes. There would be 31, 29, 31, 31, 31 duplicated changes when rattlesnake is added to positions $a-e$ or 145, 142, 145, 145, 144 minimum base changes. Taking different arrangements of the amino acids at positions 11, 12 and 15 of rattlesnake alters these values slightly but not the conclusion that *b* is the most likely position for rattlesnake. On each leg of the tree is shown the number of amino acid changes postulated to occur on that line (zero changes are not marked). Minimum base changes are shown in brackets where they are different from the amino acid changes. The total length of the branch from lamprey to the common ancestor of fish and the land vertebrates is 9 amino acid changes or 12 minimum base changes

It is desirable to check the position of rattlesnake using an ancestral sequence method. The published trees using this method have not included rattlesnake and therefore this has been done in the present work using the types of criteria used by Fitch (1971b) modified for amino acids. The procedure used has been first to use 19 vertebrate sequences and to determine the tree that required fewest amino acid changes and then to add rattlesnake to each of the 37 possible positions. In practice several variations within either the mammals or fish are equally possible and one of these best trees without rattlesnake is shown in Fig. 2. Rattlesnake is then added to all positions on Fig. 2 but only the positions marked $a-e$ are considered here. The most probable place for rattlesnake to join is on segment *b* (which

is equivalent to Fig. 1C) which is the position most consistent with the paleontological evidence. Ancestral sequence methods usually give equal weighting to each amino acid irrespective of how many base changes are required for each change. It makes no difference to the placement of rattlesnake if amino acid substitutions requiring more than one base change [and probably (Fitch, 1971a) occurring by two sequential changes] are counted as double changes to give "minimum base changes". It is possible that this tree has not been tried by other techniques or that it is caused by using different criteria for selecting the most likely tree. This latter possibility is being investigated in a separate study.

A major problem in determining phylogenetic trees is correcting the observed differences between species for repeated or reverse mutations (Ohta and Kimura, 1971; Dickerson, 1971; Holmquist, 1972). It may not be readily apparent that ancestral sequence methods automatically introduces some correction and gives a method for identifying some of the repeated or reverse mutations. The corrections can be seen by adding up all the lengths between a pair of species of Fig. 2 and comparing this with the observed (Dayhoff, 1972) differences for the same pair of species. The ancestral sequence methods do have a limitation that these corrections may be too small in long unbranched legs of the tree. However they do have the important advantage in that they use all the information from the protein sequences and not just the number of differences between proteins. The method is not dependent on the assumption (the additivity hypothesis) that the number of differences between every pair of present day proteins is a true reflection of their time and rate of divergence. For example an examination of the present distance between rattlesnake and the two primates (man and monkey) markedly underestimates the sum of the lengths of the tree (Fig. 2). This is true even if the observed difference (Dayhoff, 1972) is corrected for masked changes by *e.g.* the formula of Dickerson (1972). This could imply that there had been parallel or even convergent evolution in the cytochrome *c* on the two lines of descent.

It is not possible to decide finally between alternatives A to D of Fig. 1 since the two cytochrome *c* methods do not agree and the paleontological evidence is incomplete. However, a different calculation is possible. The trees *a* — *e* (Fig. 2) require 14–17 amino acid changes for the rattlesnake line since the time of divergence. The most probable tree includes rattlesnake at position *b* and requires a minimum of 17 amino acid substitutions on the rattlesnake line of which 5 would be double changes. The tree also requires one amino acid change between the ancestor of birds and the divergence of rattlesnake and two amino acid changes between the common ancestor of birds and snake and the divergence of turtle. From these data and from Fig. 2 it is possible to estimate both average rate of evolution of cytochrome *c* and the standard deviation (Table 1). This method is derived from the

Table 1. An estimate of the average rate and observed variability in the rate of cytochrome *c* evolution in vertebrates. The species chosen for comparison are selected so that no length of the tree in Fig. 2 is used more than once. For the first fourteen species the total number of changes has been counted from the common ancestor of the first and second species, third and fourth species, etc. and are given as both amino acid changes (column 1) and minimum base changes (column 2). The times T since divergence are taken from Ohta and Kimura (1971) and Dickerson (1971) except for the time of divergence of carp and tuna for which an estimate of 100 million years has been used (Young, 1962; Olson, 1971; Romer, 1966). The rates for both amino acid changes and minimum base changes are given for 100 million year periods (columns 4 and 5). The mean and standard deviations are given

	Amino acid Changes (*AAC*)	Minimum base Changes (*MBC*)	$T \times 10^{-8}$ (years)	Rate *AAC*/ 10^8 years	Rate *MBC*/ 10^8 years
Dogfish	11	14	3.75	2.93	3.73
Frog	14	21	3.75	3.73	5.60
Carp	1	2	1.0	1.0	2.0
Tuna	7	10	1.0	7.0	10.0
Turtle	4	4	3.0	1.33	1.33
Kangaroo	6	6	3.0	2.0	2.0
Snake	17	22	2.5	6.8	8.8
Duck	3	3	2.5	1.2	1.2
Human	8	9	0.8	10.0	11.25
Rabbit	1	2	0.8	1.25	2.5
Dog	2	2	0.6	3.33	3.33
Pig	1	1	0.6	1.67	1.67
Horse	4	6	0.9	4.44	6.67
Whale	1	2	0.9	1.11	2.22
Monkey	0	0	0.4	0	0
				$\bar{x} = 3.19$	$\bar{x} = 4.15$
				$s = 2.84$	$s = 3.50$

method of Ohta and Kimura (1971) but the pairs of species chosen in that reference would result in some lengths of Fig. 2 being used more than once and thus strictly speaking the values obtained are not independent of each other. This does not seem to give a very large effect and it is quite possible that better trees found in the future may reqiure different estimates from the one used here. Using Ohta and Kimura's species with the present data, the values for amino acid changes are $\bar{x} = 3.41$ and $s = 2.76$, and for minimum base changes $\bar{x} = 4.45$ and $s = 3.42$.

The rattlesnake line has only the fourth fastest rate of evolution and the rattlesnake values are approximately 1.3 standard deviations above the mean value. The fastest three are the primate line, the tuna line and the line from the divergence of land vertebrates down to time of divergence of mammals from birds and reptiles (see Fig. 2). This line has 10 changes in over an estimated period of 75 million years (375–300 million years). In Table 1 only results which end in a present day species have been used

because the error in estimating the time period is markedly increased when two quantities (each with a sizable error) are subtracted. The results obtained with the primate, tuna and rattlesnake are not markedly changed by small variations in the tree *e.g.* transferring rabbit to the non-primate mammals. However, small variations, such as joining land vertebrates directly with (carp, tuna, bonito) do give rather different lengths in the early part of the tree. At this time it is not possible to reject the null hypothesis that the rate of evolution of rattlesnake cytochrome *c* is not different from the average vertebrate rate. The cytochrome *c* data also gives evidence against the suggestion of Barnabas, Goodman and Moore (1972), that the rate of molecular evolution has slowed down particularly among the primates. This is a generalization based largely on a comparison of the rate of evolution of alpha hemoglobin among mammals. The slower rate of evolution of alpha hemoglobin among primates is consistent with the views of *e.g.* Dickerson (1971) because in primates any mutant on the alpha hemoglobin chain must not have an adverse effect on the beta, gamma or delta chains and not just the beta and gamma chains. But with cytochrome *c* the number of changes on the primate line shown both in Fig. 2 and by other workers (Fitch, 1971; Dayhoff, 1972; Dickerson, 1971) is larger than for other mammals. As such this is evidence against a general slowing down of the rate of molecular evolution in the primates although there still could be variations in rate along the primate line.

Discussion

Although these results (Table 1 and Fig. 2) argue against considering rattlesnake as showing a species specific effect, they do show a surprising amount of variation. Ohta and Kimura (1971) have shown that this variation is greater than can be expected by chance. The explanation could come from the work of Fitch (1971) and other workers who have shown that only 4–10% of amino acid positions (the covarions) are able to be changed at any one time although the position of covarions may change after amino acid substitutions. The rate of amino acid substitution would be expected to be proportional to both the number of covarions and, for each covarion, to the number of permissible amino acid substitutions that require only one base change (the size of the covarions). This can be illustrated more clearly as follows. In the absence of selection, the rate of mutant substitution (k) is equal to the mutation rate (v) (Kimura and Ohta, 1971)

$$i.e.\ k = v \tag{1}$$

The rate of amino acid substitution for a given protein (k_{prot}) would be expected to be

$$k_{\text{prot}} = v \cdot m \tag{2}$$

where m is the number of mutations of the existing base sequence that will result in permissible amino acid substitutions. For a strict interpretation of the evolutionary clock hypothesis it would be necessary to assume that m is constant and this is presumably the model tested by Ohta and Kimura (1971).

However it is possible that the number of covarions may undergo slight fluctuations and that the number of permissible substitutions for each covarion will vary. A more realistic model may be

$$k_{\text{prot}} = v \cdot \sum^{n} (q_i - p_i) \tag{3}$$

where n is the number of covarions, q is the number codons that would result in permissible substitutions for a given covarion and p is the number of these that cannot be reached by a one base change from the existing codon. Some simple examples can be used to illustrate this source of variability. If a given amino acid position could be one of the basic amino acids His, Lys, Arg, then there could be either zero or one possible substitutions. If the initial codon was *e.g.* CGG (Arg) then it cannot mutate to either His or Lys by a one base change but *e.g.* AGG (Arg) could mutate to AAG (Lys) but not to His. Another example would be if the same three amino acids (Arg, His, Lys) plus glutamine (Gln) were equally effective. There could be one (*e.g.* CGG), two (*e.g.* AAA) or three (*e.g.* CAA) possible amino acid substitutions depending on the initial codon. It is necessary to consider all 27 ($3 \times 3 \times 3$) possible mutations from any triplet. In the case of the last mentioned codon (CAA) there are two possibilities (CAU, CAC) giving histidine, one each for lysine (AAA) and arginine (CGA) and one possibility for a change but still staying as glutamine (CAG).

It would appear unlikely that there would be any mechanism that would result in the values of n, p and q in Eq. (3) being regulated so that the value of m in Eq. (2) remained at a constant value. It would seem possible for the rate of amino acid substitution to vary at different times and on different lines of descent but still to vary about an average rate. This would not be inconsistent with the well documented observation that the average rate of amino acid substitution is maintained over long periods and on different lines of descent (Dayhoff, 1972; Dickerson, 1971; Ohta and Kimura, 1971). Although this argument would predict that fluctuations in the evolutionary rate could occur it would be preferable not to call these "species specific". This term could be interpreted to imply an effect specific to the ecology and/or physiology of the organism. The variations in rate discussed above are not dependent on any such species specific effect and any possible increase or decrease could occur in any line of descent independently of the ecology and physiology of the organism.

This paper has only considered the first and basic argument from Jukes and Holmquist. Other subsidiary arguments were put forward and can be considered briefly. It is true that there are several double changes on the rattlesnake line but when more data are examined (Fig. 2) it is apparent that other legs have a higher proportion of double changes. Again when more data are examined (Table 1) it is apparent that the authors have compared one of the faster evolving species with some of the slowest. With at least 40 cytochrome *c* sequences known it will always be possible to find some pairs out of the 780 possible pairs that will show significantly different rates, even if the whole population is from one normal distribution (pp. 226–227, Sokal and Rohlf, 1969).

The conclusion of the present work is that the basic method used to claim a species specific effect is not valid although the paper of Jukes and Holmquist (1972) when considered "in toto" does present a serious challenge to the evolutionary clock hypothesis. The rate of evolution of rattlesnake cytochrome *c* is greater than average and has been maintained over a long period of time. Although it is not yet possible to show that rattlesnake has a significantly faster than average rate it does seem likely that fluctuations about an average rate could occur. It is still necessary to examine more closely criteria for making the optimising phylogenetic trees and it would be most worthwhile to determine more sequences of reptiles (and complete the rattlesnake sequence) both to allow a more rigorous test of the evolutionary clock hypothesis and to reduce the uncertainty in the phylogeny of the reptiles.

Acknowledgements. I thank Drs G. Carmody and P. Penny for helpful and critical discussion.

References

Barnabas, J., Goodman, M., Moore, G. W.: J. Mol. Biol. **69**, 249 (1972)
Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M., Brown, R. H.: Proc. R. Soc. London B **181**, 441 (1972)
Cox, C. B.: British Museum (Natural History) Bulletin Geology **18**, 167 (1969)
Dayhoff, M. O.: Atlas of protein sequence and structure, Vol. 5. Washington: National Biomedical Research Foundation 1972
Dickerson, R. E.: J. Mol. Evol. **1**, 26 (1971)
Dickerson, R. E.: Sci. Am. **226**, 58 (1972)
Fitch, W. M.: J. Mol. Evol. **1**, 84 (1971a)
Fitch, W. M.: Syst. Zool. **20**, 406 (1971b)
Fitch, W. M., Margoliash, E.: Brookhaven Symp. in Biol. **21**, 217 (1968)
Harland, W. B., Gilbert Smith, A., Wilcock, B. (Eds.): The phanerozoic time scale. London: Geological Society 1964
Holmquist, R.: J. Mol. Evol. **1**, 211 (1972)
Jukes, T. H., Holmquist, R.: Sci. **177**, 530 (1972)
Kimura, M., Ohta, T.: J. Mol. Evol. **1**, 1 (1971)
Ohta, T., Kimura, M.: J. Mol. Evol. **1**, 18 (1971)

Olson, E. C.: Vertebrate paleozoology. New York: Wiley 1971
Ramshaw, J. A. M., Richardson, D. L., Meatyard, B. T., Brown, R. H., Richardson, M., Thompson, E. W., Boulter, D.: New Phytol. **71** 773 (1972)
Romer, A. S.: Vertebrate paleontology, 3rd Ed. Chicago: Univ. of Chicago Press 1966
Sokal, R. R., Rohlf, F. J.: Biometry. San Francisco: Freeman 1969
Strydom, D. J., van der Walt, S. J., Botes, D. P.: Comp. Biochem. Physiol. **43**B, 21 (1972)
York, D., Farquhar, R. M.: The earth's age and geochronology. Oxford: Pergamon 1972
Young, J. Z.: The life of vertebrates, 2nd Ed. Oxford: Clarendon Press 1962

Dr. David Penny
Department of Botany and Zoology
Massey University
Palmerston North, New Zealand