

HISTORY AND PHILOSOPHY

Open Access



Rate variation during molecular evolution: creationism and the cytochrome c molecular clock

James R. Hofmann* 

Abstract

Molecular clocks based upon amino acid sequences in proteins have played a major role in the clarification of evolutionary phylogenies. Creationist criticisms of these methods sometimes rely upon data that might initially seem to be paradoxical. For example, human cytochrome c differs from that of an alligator by 13 amino acids but differs by 14 amino acids from a much more closely related primate, *Otolemur garnettii*. The apparent anomaly is resolved by taking into consideration the variable substitution rate of cytochrome c, particularly among primates. This paper traces some of the history of extensive research into the topic of rate heterogeneity in cytochrome c including data from cytochrome c pseudogenes.

Keywords: Molecular evolution, Cytochrome c, Molecular phylogenetics, Mutation rates, Molecular clocks, Creationism

Background

The most egregious and widespread creationist misuse of cytochrome c sequence data is surely the spurious “equidistance” anomaly generated when amino acid sequences taken from several members of a large clade are compared to the sequence for a single member of an outgroup to that clade. Far from being anomalous, data of this sort were both predicted and confirmed by Emanuel Margoliash as early as 1963 (Margoliash 1963, p. 677). In spite of Margoliash’s trenchant discussion of this point and numerous subsequent readily available confirmations in prominent scientific and popular sources, the claim that the cited types of equidistance are anomalies for evolutionary theory continues to circulate in creationist

venues.¹ The topic of molecular clocks thus provides unfortunate examples of how misguided creationist arguments can proliferate by means of uncritical repetition. More subtle molecular clock issues arise from the fact that mutations can result in variable amino acid replacement rates in proteins, especially among primates. For example, in a 30 September 2014 video entry for his blog, The New Creationist, Eugene Gateley pointed out that the cytochrome c of the American alligator, *Alligator mississippiensis*, has an amino acid sequence slightly closer to that of humans than is the corresponding sequence in the cytochrome c of a primate, the bush baby *Otolemur garnettii* (Gateley 2014).

To put the paradoxical aspect of this fact in context, Fig. 1 illustrates the evolutionary consensus that among early primates two major taxa, strepsirrhines and haplorhines, diverged from each other at least 70 million years ago. Haplorhines subsequently diversified into the tarsiers, the new world monkeys, the old world monkeys, the apes, and eventually *Homo sapiens*.

*Correspondence: jhofmann@fullerton.edu;
Liberal Studies Department, California State University Fullerton,
Fullerton, CA, USA

¹ This case is discussed in detail in Hofmann (2014).

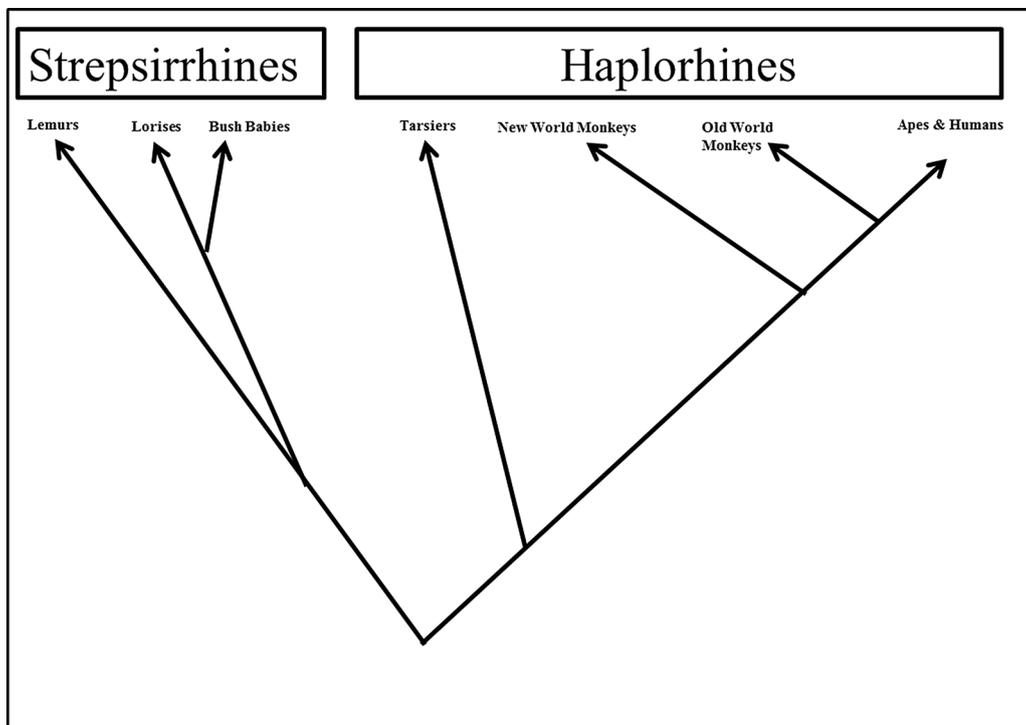


Fig. 1 A simplified phylogeny of primate evolution

Strepsirrhines also diverged into two main subgroups, lemurs and Lorisiformes. Lorisiformes in turn diversified into the lorises and the galagos, commonly referred to as bush babies. The bush baby *O. garnettii* thus is a strepsirrhine primate. Consequently, among the primates *O. garnettii* is quite distantly related to humans but is of course much more closely related to humans than is any non-primate. The data cited by Gateley thus do not appear to agree with the evolutionary consensus that the strepsirrhine primate *O. garnettii* is much more closely related to humans than alligators are. The following excerpt is a transcript of the conclusion of Gateley's recorded comments; allowance should be made for the fact that these are Gateley's spoken remarks rather than writing intended for publication.

... the question then is, why in the world would an alligator be more similar, at 87.62%, would an alligator be more similar to human than another primate at 86.67%, rounded up, you know. So, the evidence here is drastically and ridiculously contradicting the theory. And I have, this just boggles my mind how anyone could present cytochrome c as evidence for evolution in light of this evidence (Gateley 2014).

Most biochemists or molecular phylogeneticists familiar with molecular clocks would probably respond more or less flippantly that it has been known since the 1960s that cytochrome c has a variable substitution rate. Although amino acid sequence data sets for cytochrome c do have phylogenetic implications over long time periods, its relatively short amino acid sequence cannot be expected to provide precise divergence times and phylogenetic relationships in all cases. This is especially true for relatively recent and rapid processes such as the diversification of primates. Gateley is implicitly assuming that simply counting and comparing amino acid differences for three sequences is sufficient to determine the correct phylogeny for their respective species. By doing so he ignores fifty years of progress in molecular clock techniques in general and the study of mutation rates for cytochrome c in particular.²

While this response is accurate, it does not address the specific data set that Gateley cites. Why, in particular, are there fewer amino acid differences when human cytochrome c is compared to that of alligators than there

² For some of the many available summaries of these methods, see Welch and Bromham (2005), Rutschmann (2006), Lanfear et al. (2010), Bromham (2016).

are when humans are compared to the much more closely related primate *O. garrnetii*? It turns out that the example Gateley poses as if it were a new discovery of anomaly actually falls within the intersecting purviews of research areas that now are in their fifth decade. The following historical summary highlights some relevant stages in these investigations including recent analysis of cytochrome c pseudogenes. Although the stochastic nature of mutations always has to be acknowledged, a great deal of molecular evolution can be clarified, especially for a protein as thoroughly studied as cytochrome c.

Early applications and analyses of the cytochrome c molecular clock

The general idea of a molecular clock was developed by Linus Pauling and Emile Zuckerkandl shortly after the prerequisite developments in protein chemistry during the late 1950s.³ At that point it was known that each protein is constructed from a sequence of amino acids that fold into a distinctive shape required by the protein's function. Each of the 20 possible amino acid molecules consists of a carboxyl group (–COOH) opposite an amino group (–NH₂) at the other end of the molecule. In between these extremities is a so-called alpha carbon atom from which an additional side chain is attached that gives each amino acid its distinctive structure. These side chains vary considerably in size and complexity and thus can be expected to be a factor in explaining why only certain amino acids are found at crucial locations in the operative protein. When linked together in the polypeptide chain that constitutes a protein, the carboxyl group of one amino acid binds to the amino group of another with the release of a molecule of water. After this binding process, the remaining “residue” of each amino acid takes a distinctive location in the resulting polypeptide chain.

Based upon this understanding of protein structure, the initial idea of a protein molecular clock was that the number of amino acid differences found when sequencing the same protein for two different species could be used to measure the time that has elapsed since their divergence from a common ancestor. A large number of amino acid differences between two sequences would be expected to be due to a larger period of elapsed time than the time corresponding to a relatively small number of differences. Translation of a specific number of differences into an absolute measurement of time rather than a relative one requires a calibration of the clock. That is, one or more well dated events in the fossil or geological

record are used to determine the number of amino acid changes per unit of time, the rate at which a molecular clock is “ticking”.

Amino acid sequence comparisons for a specific protein can only be used as a molecular clock due to mutations in the gene coding for that protein. These mutations take place in the three-lettered DNA codons that code for the amino acids that make up the protein. The phrase “mutation rate” typically and most accurately refers to the rate at which these mutations occur. Due to redundancies in the genetic code, many of these mutations do not result in a change in amino acid. For example, codons GGT and GGC both code for the same amino acid, glycine. A “synonymous” mutation of this type from GGT to GGC would not result in one of the amino acid changes that are counted in the application of a protein molecular clock. Other mutations of course do result in a change in amino acid. For example, codon AGC codes for amino acid serine while AGA codes for arginine. The result of a “non-synonymous” mutation from AGC to AGA would be a change in amino acid that potentially would be counted in a protein molecular clock analysis. For this to be the case the relevant non-synonymous mutation must first become fixed throughout a population. Once this happens a new amino acid has been substituted in a specific location within the amino acid sequence that constitutes the protein. The rate at which these amino acid substitutions take place for a particular protein is typically referred to as the “substitution rate” or “replacement rate” for that protein. By the late 1970s researchers were also comparing DNA sequences and they often cited either mutation rates or replacement rates for the nucleotides that make up the genes that code for proteins. These rates for DNA nucleotide changes in specific genes are of course the basis for resulting amino acid substitution rates in the corresponding proteins.

Pauling and Zuckerkandl began their investigations of molecular evolution with the reasonable expectation that species with a relatively recent common ancestor should have relatively few differences when their amino acid sequences for a particular protein are compared. For example, the 104 amino acid sequence for mammalian cytochrome c is identical in humans, chimpanzees, gorillas, orangutans and gibbons. The common ancestor for these species existed so recently in the past that no cytochrome c substitutions have become fixed in any of its descendants. We are still waiting for the first “tick” of the cytochrome c clock since divergence from that common ancestor. On the other hand, species that are relatively distantly related would generally be expected to have more differences between their amino acid sequences for a particular protein. Exceptions to this general expectation can plausibly be attributed to variations

³ For historical commentary, see Dietrich (1998), Morgan (1998), Hagen (1999), Dietrich and Skipper (2007), Sommer (2008), Suarez-Diaz and Anaya-Munoz (2008), Hagen (2009, 2011), Suarez-Diaz (2014), O'Malley (2016).

in amino acid substitution rates either across species at a particular time or during an elapsed time span for particular lineages. For example, suppose a molecular clock for a particular protein has been calibrated using well established events in the fossil record. If the clock is used to study a poorly understood clade, the results will not be accurate if the protein temporarily experienced an accelerated substitution rate within that clade. Uncorrected application of this clock to two species in the clade would make them appear to be more distantly related than they actually are. That is, naïve reliance upon a temporarily accelerated molecular clock could place the common ancestor of two species farther in the past than it actually is.

Developing an explanation of an anomaly such as Gateley's by attributing it to a variable substitution rate becomes particularly apt when the relevant phylogenetic relationships can be determined with high precision independently of the molecular clock in question. In contrast to the situation during the 1960s when cytochrome c analyses were first carried out, the large number of protein and whole genome analyses and accurate fossil calibrations now at hand mean that the relevant phylogenies are sufficiently trustworthy to pinpoint the timing and nature of variable substitution rates.

One reason cytochrome c is such an extensively researched protein is due to its important function in the mitochondrial electron transport system. Eukaryotic respiratory transport systems are made up of approximately 90 proteins that collectively accomplish oxidative phosphorylation, the primary source of aerobic energy stored in ATP. Electrons are transported through four protein complexes, three of which use energy to pump protons into the intermembrane space of the mitochondrion. The potential energy in the resulting proton gradient across the membrane then drives protons back through the fifth complex of the system, ATP synthase, yielding ATP. Cytochrome c contributes to this respiratory chain by acting as an electron shuttle between Complex III (ubiquinol cytochrome c reductase) and Complex IV (cytochrome c oxidase). The initial form of this chemiosmotic theory of oxidative phosphorylation was developed by Peter Mitchell during the 1960s, the same decade in which the genetic code linking DNA codons to amino acids was deciphered.⁴

Early investigations of cytochrome c substitution rates were linked both to its molecular structure and to the role of specific amino acids in cytochrome c function. For example, throughout 1963, as amino acid sequences for cytochrome c became available for an increasing number

of species, comparisons showed that some residues vary far less frequently than others.⁵ Meanwhile, work had also begun on the tentative construction of phylogenetic trees using cytochrome c amino acid sequences. Richard Eck and Margaret Dayhoff published some of the earliest of these trees in their 1966 edition of the *Atlas of Protein Sequence and Structure* (Eck and Dayhoff 1966). Their cytochrome c based phylogeny relied upon an estimate of what the tree would be for a minimum number of amino acid substitutions. At this point they did not attempt to incorporate complications resulting from substitution rate variability. That step was taken by Margoliash and Walter Fitch who used cytochrome c data for a 1967 publication that Francisco Ayala would later refer to as "the founding document of molecular phylogenetics".⁶ Variations in the cytochrome c substitution rate were now estimated quantitatively. Fitch and Margoliash constructed phylogenetic trees based upon "mutation distances" between the cytochrome c genes for any two species. These were calculated by determining the minimum number of nucleotide replacements that would result in the transformation of the cytochrome c amino acid sequence for one species into that of another. They then argued that the most likely phylogenetic tree would be the one that minimized the composite mutation distances consistent with the amino acid sequence data.⁷ They realized that their results called attention to some lineages as particularly prone to substitution rate variation.

Thus the method indicates those lines in which the gene has undergone the more rapid changes. For example, from the point at which the primates separate from the other mammals, there are, on the average, 7.5 mutations in the descent of the former and 5.8 in that of the latter, indicating that the change in the cytochrome c gene has been much more rapid in the descent of the primates than in that of the other mammals. (Fitch and Margoliash 1967, p. 283).

Fitch and Margolis made similar comments in later publications in 1968.⁸ Further exploration of the topic

⁴ For detailed historical analyses of Mitchell's research, see Prebble and Weber (2003), Weber and Prebble (2006).

⁵ See Margoliash (1963), Smith and Margoliash (1964, p. 1244). For all seven sequences available in 1964, cysteine was found to be an invariant amino acid residue at positions 14 and 17 where bonds form to the central heme complex of the molecule. On the other hand, at position 89 six different residues were found. By 1999 113 eukaryotic cytochrome c protein sequences had been catalogued; see Banci et al. (1999). In 2013 a total of 285 sequences were catalogued and only cysteine at position 17 was known to be entirely invariant in all these species; see Zaidi et al. (2014, p. 232).

⁶ Fitch and Margoliash (1967). For Ayala's comment, see Ayala (2011, p. 5).

⁷ For historical commentary, see Felsenstein (2001) and (2004, pp. 130–133).

⁸ For examples, see Margoliash and Fitch (1968, p. 370), Margoliash et al. (1968, p. 271).

appealed to Richard Dickerson who was intrigued by the possibility that amino acid sequence comparisons might inform his primary interest in cytochrome c molecular structure and function.

Structural studies of cytochrome c

Following his initial x-ray crystallographic analysis of cytochrome c structure during the 1960s, Dickerson collaborated with illustrator Irving Geis to produce several valuable popularizations of new developments in protein biochemistry. A 1972 essay for *Scientific American* included illustrations in which Geis provided schematic representations of how the amino acid sequence of cytochrome c is coiled around the heme complex with its central iron atom (Dickerson 1972). Each amino acid residue was shown schematically as a single ball representing the alpha carbon atom from which an additional side chain would be attached in each actual amino acid structure. A point of emphasis for Dickerson was that although genetic mutation is a stochastic process, the resulting changes in amino acids are not all equally acceptable if the protein is to function properly. For example, the glycine amino acid residues at positions 6, 29, 34, 41, and 84 are located in tight corners of the cytochrome c structure where there is no room for a long side chain. Since glycine is unique in having only a single hydrogen atom as its side chain, it makes good structural sense that it is usually found in these locations. In Geis's 1972 illustration shown in Fig. 2, all 104 mammalian cytochrome c amino acids are enumerated and the 35 invariant residues known at that time are labelled using their abbreviations.

The only side chains shown are for those residues that attach to the heme, residues 14, 17, 18 and 80. The invariance of these residues for all the sequences available during the 1970s thus could plausibly be attributed to their crucial role in binding to the heme via their distinctive side chains. On the other hand, other residues such as 89 were noted to be highly variable. Residues 44 and 89 in fact will turn out to be relevant to the example raised by creationist Eugene Gateley.

Although Dickerson himself emphasized very long-term averages in substitution rates, his structural studies of cytochrome c during the 1970s coincided with much more extensive research that indicated rate variation. As Walter Fitch and Margoliash had done during the 1960s, but now with access to much more advanced statistical methods, Fitch and Charles Langley, as well as Morris Goodman and his colleagues at Wayne State, constructed increasingly detailed phylogenetic trees and then used the nodes of these trees to compare substitution rates along particular evolutionary branches. In general, the central method here was to construct the phylogenetic

tree that minimized the number of mutations compatible with the sequence data. Additional statistical factors were then introduced to compensate for gene duplications and multiple mutations at the same site, including back-mutations. Once such a tree was constructed, the number of substitutions along various branches leading to extant species could be compared. Langley and Fitch also published a series of studies in which they used expanded maximum likelihood procedures to argue for variation in the cytochrome c substitution rate.⁹ A typical conclusion drawn from their research was that "It is quite clear that the hypothesis of overall constant evolutionary rate for each protein or even overall constancy for this group of proteins as a unit must be rejected" (Langley and Fitch 1974, p. 169). Similarly, in 1976, when Goodman published a study of vertebrates with G. William Moore, Richard Holmquist, and several other coauthors, he and his colleagues could bluntly state that "Non-uniform rather than uniform rates characterize cytochrome c evolution".¹⁰ Margoliash was thoroughly convinced by these arguments, as he made clear in 1976.

Suffice it to point out that the much more precise recent study of statistical phylogenetic trees based on amino acid sequences show that the rate of evolutionary change in cytochrome c is not constant either in a single line of descent during different evolutionary intervals, or in separate lines of descent in the same evolutionary interval... (Margoliash et al. 1976, pp. 146–147).

The conclusion that the substitution rate for cytochrome c varies significantly over time thus was firmly in place by 1976. Furthermore, the molecular structure of the cytochrome c molecule was well enough understood to pick out some residues as particularly prone to substitution. It also had become clear that some of the most interesting periods of rate variation took place during the diversification of primates.

Cytochrome c among the primates

One of the primary reasons for Morris Goodman's research with primate cytochrome c was his interest in the relationship between molecular evolution and

⁹ For examples, see Langley and Fitch (1974), Fitch (1976), Fitch and Langley (1976a, b). Francisco Ayala summarized Fitch's work on this topic in his obituary tribute. "Walter demonstrated that the variance in the rate of molecular evolution was statistically larger than expected under the theory of the molecular clock and, thus, the underlying assumption of "neutral" replacements in DNA or proteins was not correct. However, he demonstrated that by combining the data from several genes or proteins, the average number of differences observed converged to the expected time since the divergence of the species investigated." Ayala (2011, p. 9).

¹⁰ Moore et al. (1976, p. 33).

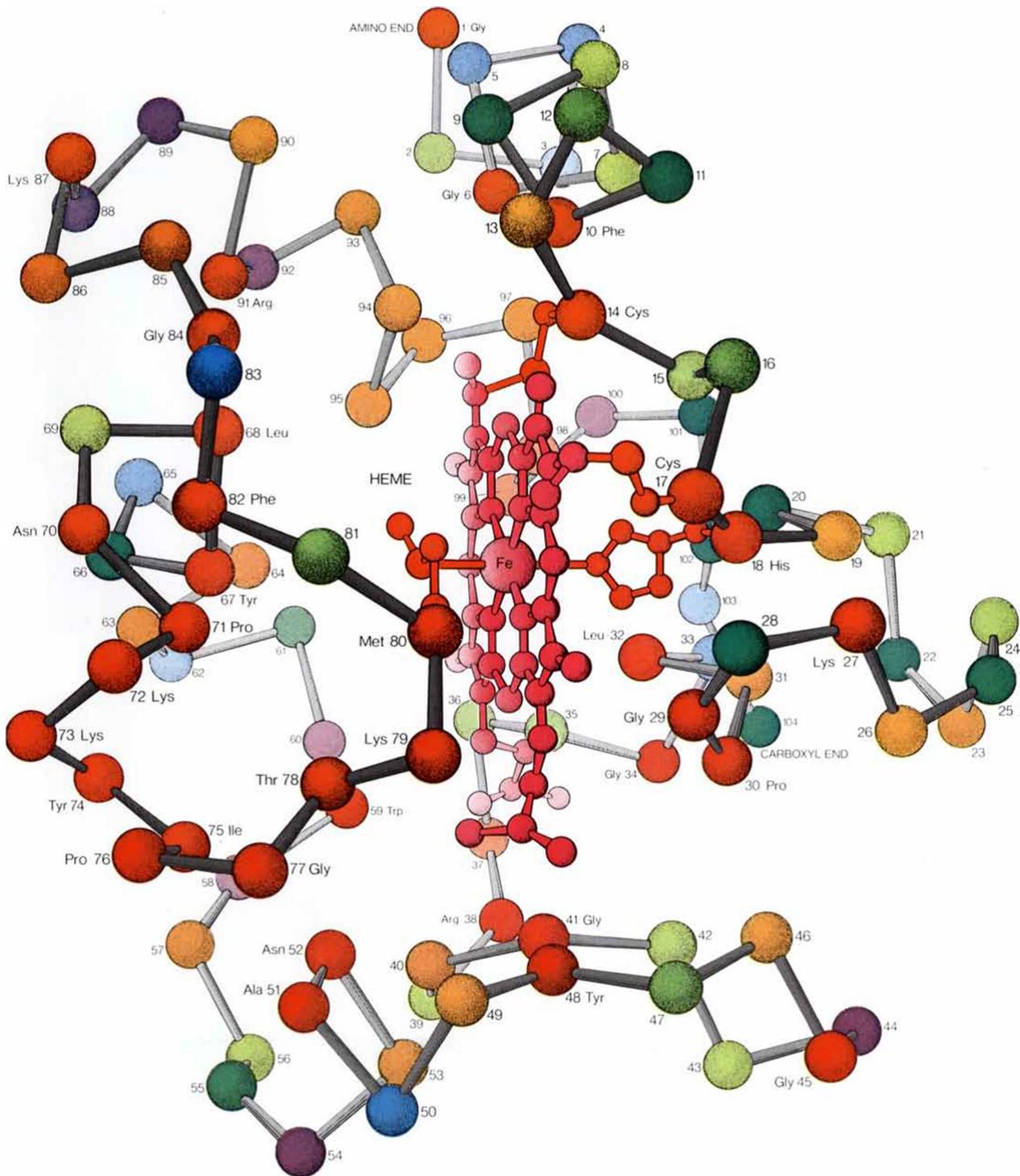


Fig. 2 A simplified folding diagram for cytochrome c. Illustration by Irving Geis in Dickerson (1972, p. 59). Highly variable residue 89 is in a peripheral location at the top of the diagram, far from the central heme. Relatively invariant residues, such as glycine 84, are labelled with their abbreviated names preceding their residue number

morphological change. During the 1970s and early 1980s Goodman and his colleagues emphasized the variability of the cytochrome c substitution rate and tried to determine whether this variability could be correlated with specific stages in primate evolution. In some particularly influential 1981–1982 publications they applied maximum parsimony methods to cytochrome c data for 87 species to construct a phylogeny from which they could compare substitution rates for a specific time period along various branches.¹¹ They expressed their results in units of “nucleotide replacements per 100 codons per 100 million years”.¹² That is, they compiled and compared data for nucleotide substitution rates rather than the resulting amino acid substitution rates. The rate of nucleotide replacements was found to peak during the period between 90 and 40 million years ago, reaching an average rate of 17.3 nucleotide replacements per 100 codons per 100 million years during that period.¹³ Since the number of amino acids in cytochrome c is 104 and thus requires 104 codons, the 17.3 replacement rate per 100 codons also corresponds to a rate of change of approximately 17.3%. This period of maximum nucleotide replacement rate and associated amino acid substitution rate stretched from the approximate date for the origins of placental mammals through the point of divergence of new world monkeys.¹⁴ Between 40 and 25 million years ago the nucleotide substitution rate dropped slightly to 12.6% and then plunged sharply to 1.9% after the 25 million year point when apes had diverged from Old World monkeys. The substitution rate thus was highest during the eras crucial for early primate radiation and then fell abruptly after 25 million years ago, a phenomenon Goodman referred to as the “hominoid slowdown”.¹⁵

Due to the inevitable incompleteness of the fossil record, particularly for primates, molecular analyses can generally be expected to give earlier divergence times than is directly supported by fossil evidence.¹⁶ The actual time of divergence of a new species from an ancestral population necessarily precedes the date assigned to the earliest relevant fossil evidence. Accurate molecular clock analysis thus can be expected to give an earlier divergence time than the date of the earliest relevant fossil. Even at present there still is some uncertainty in the

precise dating of some of the nodes in the primate phylogeny summarized in Fig. 1.¹⁷ Nevertheless, there is full agreement that the time interval Goodman highlighted between 90 and 25 million years ago includes the origin of primates, the diversification of strepsirrhines into lemurs, lorises and bush babies, and the haplorhine diversification into tarsiers, monkeys and apes. More particularly, it includes the origin of tarsiers at approximately 60–70 million years ago and the common ancestor of lorises and bush babies at approximately 40 million years ago.

Because tarsiers diverged from other haplorhines relatively early, it is customary to refer to the haplorhines other than tarsiers as anthropoids. Central to Goodman's research agenda was his argument for a link between accelerated mutation rates and functional innovations in the anthropoid molecular structure of cytochrome c. Dickerson's work was helpful in this respect since the functions of most of the 104 vertebrate cytochrome c amino acids now were at least approximately understood. Goodman analyzed the distribution of mutations over the span of amino acids in the cytochrome c sequence by distinguishing several different functional groups. His fourth group, the oxidase-reductase area of the protein, was expected to be of primary importance for phosphorylation. During the preceding decade Dickerson and Margoliash and many others had focused on 16 amino acids as probably crucial for this function; these were in positions 7, 8, 11, 12, 13, 15, 16, 19, 21, 25, 27, 72, 81, 83, 86, and 87. Five of these residues are substitution sites that distinguish human cytochrome c from that of *O. garnettii*: 11, 12, 15, 21, and 83. By 1981 Goodman and his colleagues thus had not only confirmed that cytochrome c has a variable substitution rate, but also determined that the time period for the fastest pace of change was during the early stages of primate evolution and was concentrated in residues crucial to the interaction between cytochrome c and cytochrome c oxidase during phosphorylation.

In an extensive 1990 study of cytochrome c, Geoffrey R. Moore and Graham Pettigrew summarized Goodman's results and included an illustration shown in Fig. 3 based upon one used by Goodman in 1981.¹⁸ The same two relatively recent periods of high genetic replacement rates again stand out, 25–40 million years ago and especially 40–90 million years ago, time periods that span the

¹¹ Moore et al. (1976); Baba et al. (1981), (1982); Goodman (1981).

¹² Baba et al. (1981, p. 204).

¹³ Baba et al. (1981, p. 204); and (Baba et al. 1982, pp. 20–21).

¹⁴ In the early 1980s the divergence of platyrrhines from other primates was thought to be approximately 40 million years ago; some recent estimates place it at 47 million years ago.

¹⁵ See Goodman (1985) for a summary. Goodman had argued for this idea throughout the 1960s and 1970s.

¹⁶ For a summary and many examples, see Hedges and Kumar (2003).

¹⁷ Pozzi et al. (2014) is a recent analysis that includes comparisons to several other studies carried out between 2008 and 2013.

¹⁸ Figure 6.10 from Moore and Pettigrew (1990, p. 278), based upon Figure 4 from Baba et al. (1981, p. 205). By 1986 cytochrome c had been sequenced for 92 different eukaryotic species; see Hampsey et al. (1986). Baba et al. (1982) mentions 94 species.

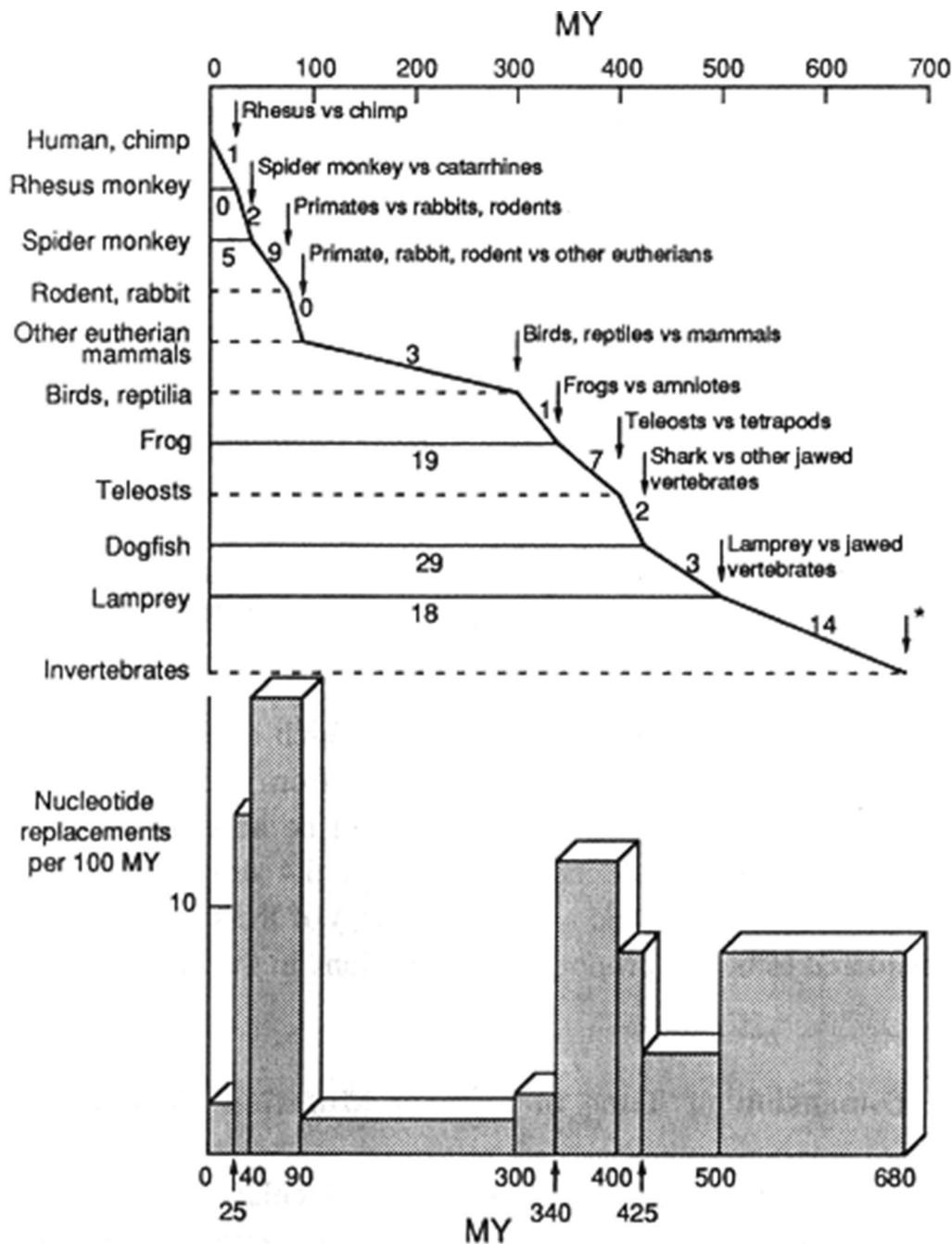


Fig. 3 The rate of change of cytochrome c. From Moore and Pettigrew (1990, p. 278)

major primate divergences, including the separation of strepsirrhines from haplorhines.

For Goodman, the recognition of mutation rate variation in cytochrome c was a preliminary motivation for further study of the causes of variation. His research thus stands in sharp contrast to more recent creationist reactions. Creationist critics typically focus on what

they interpret to be an unexpected set of data and then attempt to highlight it as a conclusive falsification of common descent. Goodman set out to see what could be learned from rate variation to understand primate evolution. In the creationist example under consideration, when a strepsirrhine primate such as the bush baby *O. garnettii* is found to have more cytochrome c amino

Table 1 Cytochrome c amino acid differences between human or ape and the bushbaby *Otolemur garnettii*

| # DIFS→ | 0 | 14 | 9 | 10 | 10 | 10 | 9 | 17 | 13 | 16 |
|---------|--------------|----------|-------------|-------------|---------|------------|--------|--------|-----------|-----------|
| AA# | Human or ape | Bushbaby | Mouse lemur | Lemur catta | Tarsier | Grey whale | Rat(s) | Rat(t) | Alligator | Bull frog |
| 1 | G | S | G | G | G | G | G | G | G | G |
| 3 | V | I | V | A | V | V | V | A | V | V |
| 11 | I | V | V | V | V | V | V | I | V | V |
| 12 | M | Q | Q | Q | Q | Q | Q | Q | Q | Q |
| 15 | S | A | A | A | A | A | A | A | A | A |
| 21 | E | D | E | E | E | E | E | E | E | E |
| 44 | P | A | A | A | A | V | A | P | P | A |
| 46 | Y | F | F | F | F | F | F | F | F | F |
| 50 | A | D | D | D | D | D | D | D | E | D |
| 58 | I | T | T | T | T | T | T | I | T | T |
| 83 | V | A | A | A | A | A | A | A | A | A |
| 85 | I | V | I | I | I | I | I | I | I | I |
| 89 | E | G | G | G | G | G | G | S | P | G |
| 96 | A | D | A | A | A | A | A | Q | A | A |

acid differences when compared to humans than an alligator does, this might suggest several questions for further research. Does other evidence exist that implies an increased or decreased mutation rate or substitution rate in one of the relevant lineages? Is this rate variation linked with similar variations in other proteins that share a function with cytochrome c in the respiratory sequence of electron transport? Have any of the relevant cytochrome c amino acids been found to be more subject to substitution than others, and if so, are there functional or adaptational reasons? Have episodes of relatively rapid protein evolution been correlated with morphological changes?

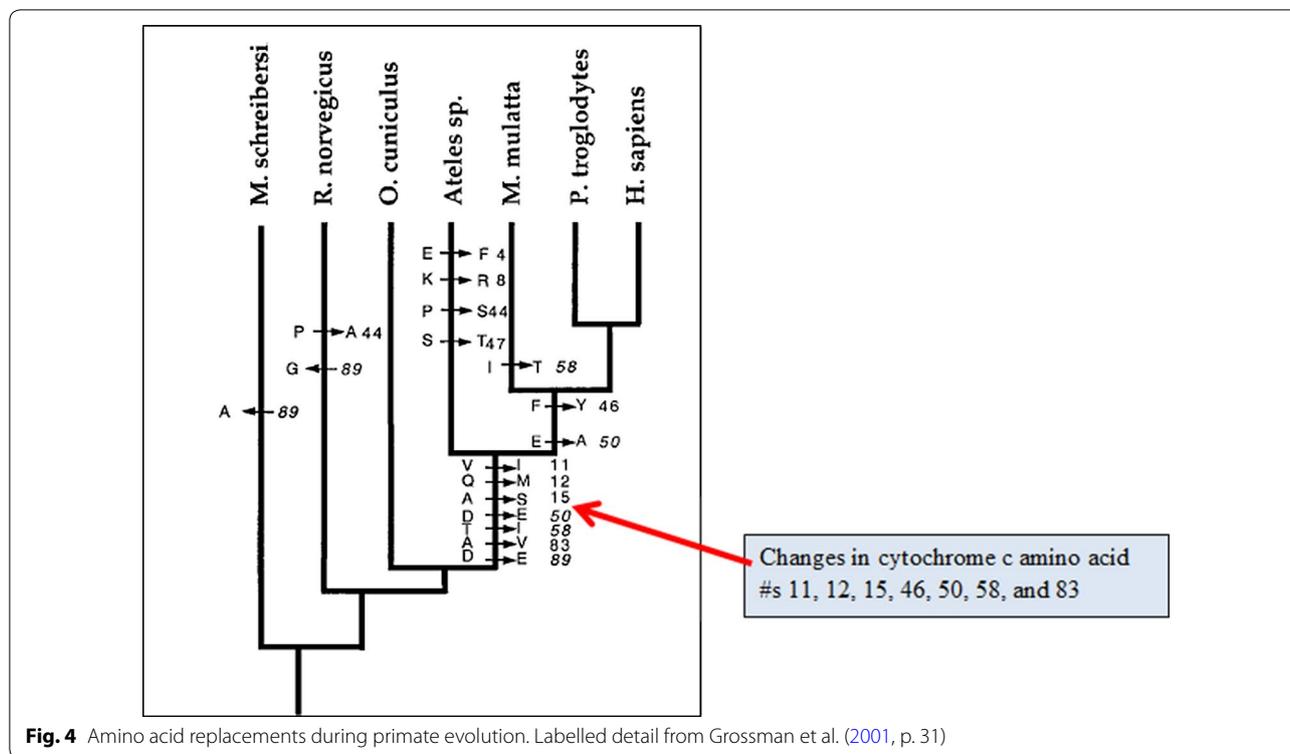
All of these questions generated productive research in the case of cytochrome c. Goodman’s group found that variable nucleotide substitution rates in the cytochrome c gene are correlated with similarly variable rates for other components of the electron transport chain, especially subunits of cytochrome c oxidase that come into direct interaction with cytochrome c during oxidative phosphorylation.¹⁹ In a 2004 review article they emphasized how the increased substitution rate in COX4-1, a sub-unit in cytochrome c oxidase, was correlated with that of cytochrome c during the same two time periods, 25–40

million years ago and 40–90 million years ago (Grossman et al. 2004). The substitution rates thus increase in multiple proteins in the electron transport chain following the divergence of anthropoid primates from tarsiers. In this analysis there was no particular reason to call attention to a specific strepsirrhine primate such as *O. garnettii*. However, if we do look at the relevant data for that species the results are quite in keeping with Goodman’s more general conclusions.

Bush babies, *Homo sapiens*, and alligators

Table 1 shows a correlated comparison of the 14 human or ape cytochrome c amino acid residues that differ from those of the bushbaby *O. garnettii*. Recall that the entire 104 amino acid sequence for cytochrome c is identical in *Homo sapiens* and all the apes. In Table 1 the residues at the 14 locations that distinguish humans and apes from *O. garnettii* are also compared to those of two lemur species, a tarsier, and several non-primate vertebrates: gray whale, rat, alligator, and bullfrog. The rat cytochrome c comes in two forms, one found in somatic cells, rat(s), and the other found exclusively in sperm cells, rat(t), and only expressed during spermatogenesis. The top row of the table shows the total number of residue differences for each species when compared to the sequence shared by humans and apes. Throughout the table, differences in specific *O. garnettii* residues when compared to the human and ape sequence are highlighted in yellow.

¹⁹ See Andrews and Eastal (2000), Grossman et al. (2001), (2004), Wildman et al. (2002), Doan et al. (2004), Pierron et al. (2011).



Locations where species have an amino acid differing from both humans and from *O. garnettii* are shown in green.

What is the most straightforward explanation for the 14 differences between human and *O. garnettii*? First of all, it is striking that so many of these amino acids are identical in most of the species listed except for humans and apes. Seven of the 14 differences between humans and *O. garnettii*, residues 11, 12, 15, 46, 50, 58, and 83, apparently involve mutations in the relatively recent anthropoid lineage that leads to monkeys, apes and humans after their divergence from tarsiers and long after their earlier divergence from strepsirrhines such as *O. garnettii* and the lemurs.²⁰ Five of the remaining amino acid differences appear to have happened along the divergent branch leading to the strepsirrhine bush baby *O. garnettii* (residues 1, 3, 21, 85, and 96). Residues 44 and 89 have apparently undergone multiple substitutions resulting in differences not only between humans and *O. garnettii* but with the other listed species as well. This is not surprising since residues 44 and 89 were discovered by Dickerson to be located far from the heme

core of the cytochrome c molecule and thus allow a high degree of variability.

These data are in keeping with the temporarily accelerated substitution rate thoroughly documented since the early 1980s. Goodman and his colleagues had in fact included a cytochrome c analysis in one of their 2001 studies of the evolution of the electron transfer complex (Grossman et al. 2001). Figure 4 highlights some details from their illustration of the extensive amino acid replacements occurring along the Catarrhine stem after the divergence of both *Rattus norvegicus* (brown rat) and *Oryctolagus cuniculus* (European rabbit) and prior to the divergence of Old World monkeys such as *Ateles* sp. (spider monkey) 25 million years ago. Along with highly variable residue 89, the figure labels precisely those seven amino acid changes that stand out from a straightforward perusal of the data (amino acid #s 11, 12, 15, 46, 50, 58, and 83). Additional changes in the highly variable residues 44 and 89 are also indicated.

Although data for *O. garnettii* are not shown in this diagram, as a strepsirrhine primate it diverged from the other primates prior to the highlighted changes in the Catarrhine stem leading to monkeys and apes as well as *Homo sapiens*.

Seven of the 14 amino acid differences between human and *O. garnettii* thus are accounted for by recent changes in the anthropoid lineage, five can be attributed to the

²⁰ One exception for residue position 58 is an old world monkey, the hamadryas baboon, which also has T at position 58. This is one of three differences it has when compared to humans. The others are at positions 4 and 33.

Table 2 Amino acid differences between human or ape and alligator cytochrome c

| AA# | Human or ape | Alligator | Mouse lemur | Lemur catta | Tarsier | Grey whale | Kangaroo | Hippo | Horse | Bushbaby | Bull frog | Tuna |
|-----|--------------|-----------|-------------|-------------|---------|------------|----------|-------|-------|----------|-----------|------|
| 11 | I | V | V | V | V | V | V | V | V | V | V | V |
| 12 | M | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q | Q |
| 15 | S | A | A | A | A | A | A | A | A | A | A | A |
| 36 | F | I | F | F | F | F | F | F | F | F | I | F |
| 46 | Y | F | F | F | F | F | F | F | F | F | F | Y |
| 50 | A | E | D | D | D | D | D | D | D | D | D | D |
| 58 | I | T | T | T | T | T | I | T | T | T | T | V |
| 62 | D | E | D | D | D | E | D | E | E | D | D | N |
| 83 | V | A | A | A | A | A | A | A | A | A | A | A |
| 89 | E | P | G | G | G | G | G | G | T | G | G | G |
| 100 | K | E | K | K | K | K | K | Q | K | K | S | S |
| 103 | N | S | N | N | N | N | N | N | N | N | S | S |
| 104 | E | N | E | E | E | E | E | E | E | E | K | |

Table 3 Amino acid differences for alligator and primates

| | Alligator | <i>Microcebus coquereli</i> (mouse lemur) | <i>tarsius bancanus</i> (Western tarsier) | <i>Lemur catta</i> (ring-tailed lemur) | <i>Otolemur garnettii</i> (bush-baby) | Human and ape |
|----------|-----------|--|--|---|--|---------------|
| AA DIFS→ | 0 | 8 | 9 | 9 | 13 | 13 |

strepsirrhine lineage leading to *O. garnettii*, and the remaining two, 44 and 89, have been subject to multiple substitutions. Substitution rates among some strepsirrhine lineages have more recently been found to be generally very high, even compared to other primates. These conclusions are of course based on much more thorough sequencing techniques than earlier ones that relied simply upon individual proteins (Eizirik et al. 2004, pp. 54–55). The upshot of these and many other studies is that instead of using cytochrome c simplistically as a molecular clock assumed to have a fixed substitution rate, other more reliable timing mechanisms have been used to link the variable substitution rate of cytochrome c to its structure and function and to particular episodes in primate evolution.

Similar conclusions can be drawn from the data for a comparison of alligators and humans shown in Table 2. Of the thirteen amino acid differences between alligators and *Homo sapiens*, six can be assigned solely to the anthropoid lineage (11, 12, 15, 46, 58, and 83).

Two others apparently involve substitutions in both anthropoids and crocodylians (50 and 89), and five are found only in the crocodylian lineage (36, 62, 100, 103, and 104). One question that these data prompt is why

only a total of seven amino acid replacements have taken place along the long crocodylian lineage in contrast to the eight assigned to a much shorter time period within the primate lineage. A reasonable place to look for an explanation would be to see what the mutation rate is in the crocodylian lineage. It turns out that in contrast to primates, the crocodylian lineage has an unusually low genetic substitution rate. In their 2014 study using whole genome-alignments, Richard Green and colleagues found that alligators and crocodiles have “exceptionally low rates of evolution relative to mammals” (Green et al. 2014, 1254449-3). As a result, it is not surprising that only seven crocodylian cytochrome c amino acid replacements have contributed to the difference between human and alligator cytochrome c. Eight *O. garnettii* substitutions took place during a much shorter time period.

One more aspect of the alligator, *O. garnettii*, and human cytochrome c data is worth mentioning. As shown in Table 3, human and *O. garnettii* cytochrome c sequences both differ from alligator cytochrome c by 13 amino acids. Of the 13 differences between human and alligator sequences, six are at amino acids where all the other primates listed have the same amino acid as alligator (amino acid #s 11, 12, 15, 46, 58, 83). As we have

seen, all of these six differences have arisen in the anthropoid lineage. Five of the remaining seven differences are shared by humans and the other primates listed (amino acid #s 36, 62, 100, 103, and 104). The data thus are very much as would be expected from accelerated mutation and substitution rates within primate lineages.

In sharp contrast to the multi-faceted investigation of molecular evolution by the scientific community, creationist responses to cytochrome c data demonstrate quite a different attitude. A common reaction is to simply use intuitively unexpected cytochrome c sequence data for specific primates as a reason to categorically reject amino acid sequence data as evidence for common descent. Eugene Gateley presents his examples as if they are recent discoveries, even giving the impression that he might be the first to have noticed them. Data sets of this type have in fact been subject to interesting research for decades, research that securely explains them as consequences of variation in the rate of amino acid replacement. The apparent anomaly generated by the 14 differences between human and *O. garnettii* cytochrome c amino acid sequences thus is resolved as part of a more general analysis of the variable rates of molecular evolution in cytochrome c.

Human cytochrome c pseudogenes

Interesting additional confirmation of the variable substitution rate in the evolutionary history of cytochrome c comes from its numerous pseudogenes. In general, pseudogenes are versions of a gene that no longer carry out that gene's initial function. In some cases unitary pseudogenes are the direct remains of a gene that has become dysfunctional due to mutations. In other cases a gene has undergone duplication and one copy has mutated and become a pseudogene. In still other cases a processed pseudogene is the result of transcription and retrotransposition, that is, reinsertion of a nucleotide sequence back into the genome after being transcribed, stripped of introns, and then left without a promoter to generate subsequent transcription. Processed pseudogenes thus are relatively easy to identify due to their lack of introns.²¹

The fact that human cytochrome c has a large number of processed pseudogenes attracted research interest during the 1980s.²² In humans the functioning gene for cytochrome c is located on chromosome 7 and has two

introns. By 2003 Zhaolei Zhang and Mark Gerstein had identified 49 cytochrome c pseudogenes distributed over 18 different human chromosomes (Zhang and Gerstein 2003). They called particular attention to nine highly variable residues (11, 12, 15, 44, 46, 50, 58, 83, 89), all of which are among the 14 residues that distinguish human cytochrome c from that of *O. garnettii*. As we have seen, all of these substitutions, except for the highly variable sites 44 and 89 have been attributed to mutations that took place solely in the anthropoid lineage long after anthropoid divergence from strepsirrhines such as *O. garnettii*. Zhang and Gerstein followed a prior protocol in distinguishing between two sets of cytochrome c pseudogenes. The four pseudogenes in class 1 (ψ 15, ψ 21, ψ 45 and ψ 46) all code for sequences that have a high degree of similarity to the functional human cytochrome c. This implies that the cytochrome c gene experienced a period of significant mutation relatively recently in the anthropoid lineage that gave rise to the four pseudogenes in class 1 only after these mutations. Table 4 shows class 1 pseudogene data for all 14 of the amino acid differences for *O. garnettii* compared to humans or apes along with the somatically expressed rat cytochrome c, rat(s). Differences from human cytochrome c are shown in yellow.

These data contributed to the 2001 conclusion by Goodman's research group that the class 1 pseudogenes originated during a period of accelerated cytochrome c substitution rate between 40 and 25 million years ago.²³ As we have seen, they assigned substitutions in amino acid #s 11, 12, 15, 46, 50, 58, and 83 solely to the anthropoid lineage. The pseudogenes in class 1 all came about after these substitutions and preserved them in all but a very few residues.

Secondly, Zhang and Gerstein placed the remaining 45 of the total 49 pseudogenes in a set labelled class 2. They noted that the amino acid sequences coded for by these pseudogenes bear very few identities to human cytochrome c at highly variable locations such as 11, 12, 15, 44, 46, 50, 58, 83, and 89. The most straightforward interpretation of the data is that the 45 members of class 2 are relatively old pseudogenes compared to class 1. Zhang and Gerstein used the known age of retrotransposon insertions to estimate the age of the oldest class 2 pseudogene to be at least 80 million years. As a result of their age, and in contrast to the pseudogenes in class 1, pseudogenes in class 2 should have a relatively high degree of correlated similarities or dissimilarities to both the *O. garnettii* gene and the human gene at amino acid positions that distinguish *O. garnettii* from humans and apes.

²¹ See Zhang and Zheng (2014).

²² See Wu et al. (1986), Evans and Scarpulla (1988), Virbasius and Scarpulla (1988), Zhang and Gerstein (2003, 2004), Mills (1991), Zhang et al. (2003). Due to the relatively small data sets available during the 1980s, early studies were carried out in conjunction with comparisons to genes and pseudogenes in rodents. Recent analysis based upon both fossil evidence and molecular clocks place the origin of rodents at approximately 60 million years ago during the period of very high cytochrome c substitution rate. See Wu et al. (2012).

²³ Grossman et al. (2001, p. 31).

Table 4 Class 1 human cytochrome

| Amino acid # | 1 | 3 | 11 | 12 | 15 | 21 | 44 | 46 | 50 | 58 | 83 | 85 | 89 | 96 |
|---------------------------|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| Human/ape | G | V | I | M | S | E | P | Y | A | I | V | I | E | A |
| ψ15 | G | V | I | M | S | E | P | H | A | I | V | I | E | A |
| ψ21 | G | V | I | M | S | E | P | Y | A | I | V | I | E | A |
| ψ45 | G | V | I | M | S | E | A | Y | V | T | V | I | E | A |
| ψ46 | | | | | | | P | Y | A | I | V | I | E | A |
| Rat(s) | G | V | V | Q | A | E | A | F | D | T | A | I | G | A |
| <i>Otolemur garnettii</i> | S | I | V | Q | A | D | A | F | D | T | A | V | G | D |

For example, as illustrated in Fig. 5, at positions 1, 3, 21, 85, and 96 we would expect to see differences between class 2 pseudogenes and *O. garnettii* but similarities to humans. This is because mutations took place for these residues only in the strepsirrhine lineage leading to *O. garnettii* but not in the anthropoid lineage. On the other hand, at positions 11, 12, 15, 46, 50, 58, and 83 we should see just the opposite, namely, a similarity to *O. garnettii* and dissimilarities to humans. Mutations at these

locations took place only relatively late in the anthropoid lineage leading to humans and apes but not in the strepsirrhine lineage leading to *O. garnettii*. These expectations are summarized in Fig. 6.

The highly variable sites 44 and 89 can be expected to differ from both human and *O. garnettii* due to mutations in both the strepsirrhine and anthropoid lineages. Of course since pseudogenes generally sustain arbitrary mutations to a higher degree than do functioning genes,

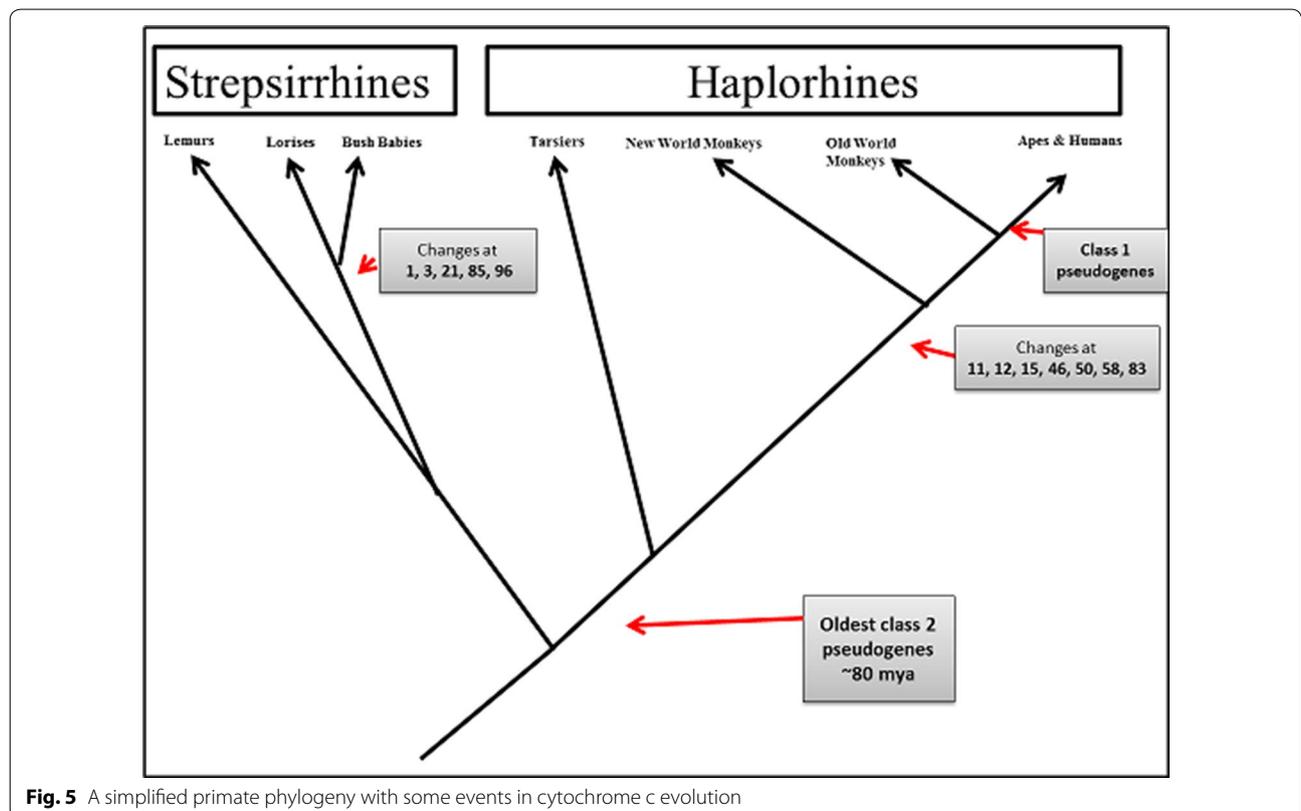
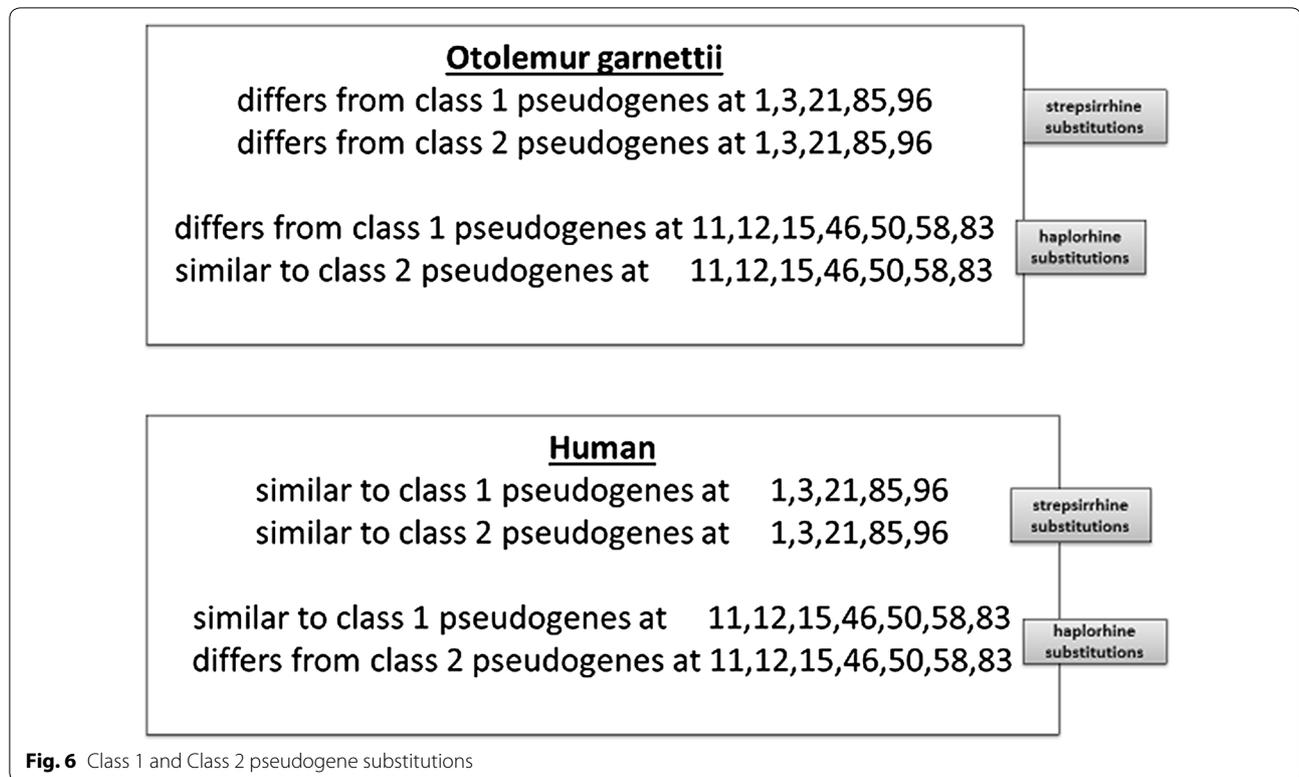


Fig. 5 A simplified primate phylogeny with some events in cytochrome c evolution



we should not expect these correlations to be without exceptions. Nevertheless, the data do generate quite striking patterns. Figure 7 shows the data with colors coordinated for class 2 pseudogene residues that match either *O. garnettii* or human cytochrome c. Residues that match neither species are shown in red.

These data thus once again confirm the conclusion that mutations at positions 11, 12, 15, 46, 50, 58, and 83 all took place in the anthropoid lineage leading to humans and that mutations in residues 1, 3, 21, 85, and 96 came about within the strepsirrhine lineage leading to *O. garnettii*. The highly variable sites 44 and 89 have undergone multiple substitutions with the result that neither human nor *O. garnettii* has very many similarities to any of the ancient class 2 pseudogenes at these locations. As Zhang and Gerstein concluded, “our findings strongly support the hypothesis that this gene has evolved at a very rapid rate in the recent human lineage” (Zhang and Gerstein 2003, p. 71). More specifically, the pseudogene data support detailed phylogenetic assignments for all the amino acid residues that distinguish human cytochrome c from that of *O. garnettii*. The cytochrome c mutation and substitution rates certainly are not claimed to be arbitrarily variable. On the contrary, specific amino acid changes can plausibly be assigned to either the anthropoid or the *O. garnettii* lineage in such a way as to be compatible with all the protein and pseudogene data.

Conclusion

The history of cytochrome c research shows that the present understanding of its mutation rate heterogeneity has progressed in conjunction with study of its molecular structure and its associated pseudogenes. The initial correlation of cytochrome c substitution rate changes with specific time intervals was gradually supplemented by sequence data from entire genomes and a multitude of other molecular clocks. Rather than simplistically using the irregular cytochrome c clock to determine precise primate phylogenies, much broader data sets have clarified interesting episodes in the variation of the cytochrome c substitution rate. Although cytochrome c no longer plays a cutting edge role in the determination of divergence dates among primates, it has played an important role in the historical development of the field. The contrast between scientific inquiry into this topic and creationist commentary is severe. When calling attention to an apparent molecular clock anomaly, it is not illuminating to simply count and compare amino acid residue differences without asking further questions. This is particularly true when many of the relevant questions have received increasingly detailed answers over several decades. Zuckerkandl and Pauling made this point over 50 years ago.

Counting numbers of differences in amino acid sequence is only one stage of the analysis, and record-

| Amino acid | 11 | 12 | 15 | 46 | 50 | 58 | 83 | Amino acid | 1 | 3 | 21 | 85 | 96 | Amino acid | 44 | 89 |
|------------|----|----|----|----|----|----|----|------------|---|---|----|----|----|------------|----|----|
| Human/ape | I | M | S | Y | A | I | V | Human/ape | G | V | E | I | A | Human/ape | P | E |
| Bushbaby | V | Q | A | F | D | T | A | Bushbaby | S | I | D | V | D | Bushbaby | A | G |
| ψ1 | V | Q | A | F | D | I | A | ψ1 | G | V | E | I | A | ψ1 | V | A |
| ψ2 | V | Q | A | F | D | T | A | ψ2 | G | I | E | I | D | ψ2 | T | _ |
| ψ3 | F | R | S | F | D | I | T | ψ3 | G | V | _ | I | A | ψ3 | /P | T |
| ψ4 | | | A | /Y | D | T | V | ψ4 | | | V | T | A | ψ4 | T | A |
| ψ5 | V | X | V | F | D | I | T | ψ5 | G | V | E | V | P | ψ5 | A | A |
| ψ6 | V | Q | A | F | D | T | A | ψ6 | G | V | _ | I | A | ψ6 | I | A |
| ψ7 | V | Q | A | F | D | T | A | ψ7 | S | V | E | I | A | ψ7 | V | A |
| ψ8 | F | T | T | S | D | T | T | ψ8 | S | V | E | I | T | ψ8 | I | V |
| ψ9 | I | Q | A | F | E | V | S | ψ9 | G | A | E | I | Q | ψ9 | P | S |
| ψ10 | V | Q | A | F | D | T | A | ψ10 | S | V | E | I | A | ψ10 | V | A |
| ψ11 | V | Q | A | F | D | T | A | ψ11 | G | A | E | I | A | ψ11 | T | A |
| ψ12 | V | Q | A | F | D | T | A | ψ12 | G | A | E | I | A | ψ12 | I | A |
| ψ13 | V | Q | A | F | D | T | A | ψ13 | G | A | E | I | A | ψ13 | I | A |
| ψ14 | F | L | V | F | D | I | I | ψ14 | | A | E | I | Q | ψ14 | P | _ |
| ψ16 | V | Q | S | P | _ | X | | ψ16 | | V | E | | | ψ16 | R | |
| ψ17 | I | Q | V | F | D | T | F | ψ17 | G | V | E | /V | T | ψ17 | I | A |
| ψ18 | V | /M | A | F | V | S | | ψ18 | | V | K | | | ψ18 | P | |
| ψ19 | I | Q | A | L | D | T | A | ψ19 | D | V | E | I | A | ψ19 | V | _ |
| ψ20 | V | Q | A | F | D | T | G | ψ20 | G | V | E | I | A | ψ20 | I | G |
| ψ22 | V | Q | V | F | D | T | A | ψ22 | G | V | E | I | A | ψ22 | P | V |
| ψ23 | V | Q | A | F | E | S | A | ψ23 | G | V | E | M | A | ψ23 | V | A |
| ψ24 | V | Q | A | F | D | T | A | ψ24 | G | V | E | I | A | ψ24 | V | A |
| ψ25 | V | Q | A | F | D | T | A | ψ25 | G | V | E | I | A | ψ25 | I | A |
| ψ26 | V | Q | A | F | D | T | A | ψ26 | G | V | E | I | A | ψ26 | V | A |
| ψ27 | T | E | A | F | D | T | A | ψ27 | G | I | E | I | A | ψ27 | I | _ |
| ψ28 | V | E | V | F | D | T | T | ψ28 | G | A | E | I | A | ψ28 | A | /A |
| ψ29 | V | Q | A | Y | D | T | A | ψ29 | G | A | Q | I | A | ψ29 | V | A |
| ψ30 | V | Q | A | | D | T | A | ψ30 | G | V | E | I | A | ψ30 | _ | T |
| ψ31 | V | Q | A | F | D | T | A | ψ31 | G | V | E | I | A | ψ31 | V | A |
| ψ32 | V | Q | A | L | E | T | A | ψ32 | G | V | E | I | A | ψ32 | I | A |
| ψ33 | V | Q | A | L | D | T | S | ψ33 | R | V | E | I | T | ψ33 | I | T |
| ψ34 | V | Q | A | F | D | T | A | ψ34 | G | V | /E | I | A | ψ34 | I | A |
| ψ35 | K | K | T | S | E | T | A | ψ35 | | | E | I | A | ψ35 | A | S |
| ψ36 | V | Q | A | F | D | T | A | ψ36 | S | A | E | I | A | ψ36 | A | T |
| ψ37 | A | Q | A | F | D | A | A | ψ37 | G | V | E | I | A | ψ37 | V | V |
| ψ38 | V | X | A | F | D | T | Y | ψ38 | G | V | G | K | | ψ38 | F | |
| ψ39 | V | Q | A | F | D | T | A | ψ39 | G | V | E | I | A | ψ39 | I | A |
| ψ40 | V | P | A | F | D | T | T | ψ40 | D | V | E | | | ψ40 | A | |
| ψ41 | V | Q | A | L | E | T | | ψ41 | G | A | E | | | ψ41 | I | |
| ψ42 | V | Q | S | F | D | A | A | ψ42 | | | E | V | | ψ42 | P | |
| ψ43 | A | Q | A | F | D | N | T | ψ43 | V | V | E | I | A | ψ43 | V | A |
| ψ44 | V | Q | A | F | D | T | S | ψ44 | G | V | E | I | A | ψ44 | V | T |
| ψ47 | | | S | | | | | ψ47 | | | E | | | ψ47 | | |
| ψ48 | | | A | /Y | D | T | A | ψ48 | | | V | T | A | ψ48 | T | A |
| ψ49 | L | Q | A | L | E | T | A | ψ49 | A | S | E | I | A | ψ49 | I | A |

Fig. 7 Class 2 pseudogenes

ing the nature of the differences is a necessary further step in the establishment of a molecular phylogeny. (Zuckerandl and Pauling 1965, pp. 137–138).

Acknowledgements

Nick Matzke provided very constructive comments on early drafts of this paper.

Competing interests

The author declares that he has no competing interests.

Received: 17 January 2016 Accepted: 13 January 2017

Published online: 06 February 2017

References

- Andrews T, Easteal S. Evolutionary rate acceleration of cytochrome c oxidase subunit I in simian primates. *J Mol Evol.* 2000;50:562–8.
- Ayala F, Walter Monroe Fitch 1929–2011 a biographical memoir. Washington, DC: National Academy of Sciences; 2011.
- Baba M, Darga L, Goodman M, Czelusniak J. Evolution of cytochrome c investigated by the maximum parsimony method. *J Mol Evol.* 1981;17:197–213.
- Baba M, Darga L, Goodman M. Recent advances in molecular evolution of the primates. In: Chiarelli AB, Corruccini RS, editors. *Advanced views in primate biology*. New York: Springer; 1982. p. 6–27.
- Banci L, Bertini I, Rosato A, Varani G. Mitochondrial cytochromes c: a comparative analysis. *J Biol Inorg Chem.* 1999;4:824–37.
- Bromham L. *An introduction to molecular evolution and phylogenetics*. Oxford: Oxford University Press; 2016.
- Dickerson R. The structure and history of an ancient protein. *Sci Am.* 1972;226(4):58–75.
- Dietrich M. Paradox and persuasion: negotiating the place of molecular evolution within evolutionary biology. *J Hist Biol.* 1998;31:85–111.
- Dietrich M, Skipper R. Manipulating underdetermination in scientific controversy: the case of the molecular clock. *Perspect Sci.* 2007;15(3):295–325.
- Doan J, Schmidt T, Wildman D, Uddin M, Goldberg M, Hüttemann M, Goodman M, Weiss M, Grossman L. Coadaptive evolution in cytochrome c oxidase: 9 of 13 subunits show accelerated rates of nonsynonymous substitution in anthropoid primates. *Mol Phylogenet Evol.* 2004;33(3):944–50.
- Eck R, Dayhoff M, editors. *Atlas of protein sequence and structure 1966*. Silver Spring: National Biomedical Research Foundation; 1966.
- Eizirik E, Murphy W, Springer M, O'Brien S. Molecular phylogeny and dating of early primate divergences. In: Ross CF, Kay RF, editors. *Anthropoid origins: new visions*. New York: Kluwer Publishers; 2004. p. 45–64.
- Evans M, Scarpulla R. The human somatic cytochrome c gene: two classes of processed pseudogenes demarcate a period of rapid molecular evolution. *Proc Natl Acad Sci.* 1988;85:9625–9.
- Felsenstein J. The troubled growth of statistical phylogenetics. *Syst Biol.* 2001;50(4):465–7.
- Felsenstein J. *Inferring phylogenies*. Sunderland: Sinauer; 2004.
- Fitch W. Molecular evolutionary clocks. In: Ayala FJ, editor. *Molecular evolution*. Sunderland: Sinauer; 1976. p. 160–78.
- Fitch W, Langley C. Protein evolution and the molecular clock. *Fed Proc.* 1976a;35(10):2092–7.
- Fitch W, Langley C. Evolutionary rates in proteins: neutral mutations and the molecular clock. In: Goodman M, Tashian R, editors. *Molecular anthropology: genes and proteins in the evolutionary ascent of the primates*. New York: Plenum; 1976b. p. 197–219.
- Fitch W, Margoliash E. Construction of phylogenetic trees. *Science.* 1967;155:279–84.
- Gateley E. Cytochrome c contradicts evolution. 2014. <http://new-creationist.blogspot.com/2014/09/cytochrome-c-contradicts-evolution.html>. Accessed 29 Sept 2016. https://www.youtube.com/watch?v=h24vYJ_0X_Y. Accessed 29 Sept 2016.
- Goodman M. Decoding the pattern of protein evolution. *Prog Biophys Mol Biol.* 1981;38:105–64.
- Goodman M. Rates of molecular evolution: the hominoid slowdown. *BioEssays.* 1985;3(1):9–14.
- Green R, et al. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science.* 2014;346(6215):1335. doi:10.1126/science.1254449.
- Grossman L, Schmidt T, Wildman D, Goodman M. Molecular evolution of aerobic energy metabolism in primates. *Mol Phylogenet Evol.* 2001;18(1):26–36.
- Grossman L, Wildman D, Schmidt T, Goodman M. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends Genet.* 2004;20(11):578–85.
- Hagen J. Naturalists, molecular biologists, and the challenges of molecular evolution. *J Hist Biol.* 1999;32:321–41.
- Hagen J. Descended from Darwin? George Gaylord Simpson, Morris Goodman, and Primate Systematics. *Trans Am Philos Soc New Ser.* 2009;99(1):93–109.
- Hagen J. The origin and early reception of sequence databases. In: Hamacher M, Eisenacher M, Stephan C, editors. *Data mining in proteomics: from standards to applications*. New York: Springer; 2011. p. 61–78.
- Hampsey D, Das G, Sherman F. Amino acid replacements in yeast iso-1-cytochrome c: comparison with the phylogenetic series and the tertiary structure of related cytochromes c. *J Biol Chem.* 1986;261(7):3259–71.
- Hedges SB, Kumar S. Genomic clocks and evolutionary timescales. *Trends Genet.* 2003;19(4):200–6.
- Hofmann J. A tale of two crocoducks: creationist misuses of molecular evolution. *Sci Educ.* 2014;23(10):2095–117.
- Lanfear R, Welch J, Bromham L. Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol Evol.* 2010;25:495–503.
- Langley C, Fitch W. An examination of the constancy of the rate of molecular evolution. *J Mol Evol.* 1974;3:161–77.
- Margoliash E. Primary structure and evolution of cytochrome c. *Proc Natl Acad Sci.* 1963;50:672–9.
- Margoliash E, Ferguson-Miller S, Brautigan D, Chaviano A. Functional basis for evolutionary change in cytochrome c structure. In: Markham R, Horne RW, editors. *Structure-function relationships of proteins: proceedings of the third John Innes symposium held in Norwich, July 1976*. Amsterdam: North-Holland; 1976. p. 145–65.
- Margoliash E, Fitch W. Evolutionary variability of cytochrome c primary structures. *Ann NY Acad Sci.* 1968;151:359–81.
- Margoliash E, Fitch W, Dickerson R. Molecular expression of evolutionary phenomena in the primary and tertiary structures of cytochrome c. *Brookhaven Symp Biol.* 1968;21(2):259–305.
- Mills G. Cytochrome c: gene structure, homology and ancestral relationships. *J Theor Biol.* 1991;152:177–90.
- Moore G, Goodman M, Callahan C, Holmquist R, Moise H. Stochastic versus augmented maximum parsimony method for estimating superimposed mutations in the divergent evolution of protein sequences. Methods tested on cytochrome c amino acid sequences. *J Mol Biol.* 1976;105(1):15–37.
- Moore G, Pettigrew G, editors. *Cytochromes c: evolutionary, structural and physicochemical aspects*. New York: Springer Verlag; 1990.
- Morgan G. Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. *J Hist Biol.* 1998;31:155–78.
- O'Malley M. Histories of molecules: reconciling the past. *Stud Hist Philos Sci.* 2016;55:69–83.
- Pierron D, Opazo J, Heiske M, Papper Z, Uddin M, Chand G, Wildman D, Romero R, Goodman M, Grossman L. Silencing, positive selection and parallel evolution: busy history of primate cytochromes c. *PLoS ONE.* 2011;6910(e26269):2016. doi:10.1371/journal.pone.0026269.
- Pozzi L, Hodgson J, Burrell A, Sterner K, Raaum R, Distell T. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Mol Phylogenet Evol.* 2014;75:165–83.
- Prebble J, Weber B. *Wandering in the gardens of the mind*. Peter Mitchell and the making of Glynn. New York: Oxford University Press; 2003.
- Rutschmann F. Molecular dating of phylogenetic trees: a brief review of current methods that estimate divergence times. *Divers Distrib.* 2006;12:35–48.

- Smith E, Margoliash E. Evolution of cytochrome c. *Fed Proc.* 1964;23:1243–7.
- Sommer M. History in the gene: negotiations between molecular and organismal anthropology. *J Hist Biol.* 2008;41:473–528.
- Suarez-Diaz E. The long and winding road of molecular data in phylogenetic analysis. *J Hist Biol.* 2014;47:443–78.
- Suarez-Diaz E, Anaya-Munoz V. History, objectivity, and the construction of molecular phylogenies. *Stud Hist Philos Biol Biomed Sci.* 2008;39:451–68.
- Virbasius J, Scarpulla R. Structure and expression of rodent genes encoding the testis-specific cytochrome c. *J Biol Chem.* 1988;263(14):6791–6.
- Weber B, Prebble J. An issue of originality and priority: the correspondence and theories of oxidative phosphorylation of Peter Mitchell and Robert JP Williams, 1961–1980. *J Hist Biol.* 2006;39:125–63.
- Welch J, Bromham L. Molecular dating when rates vary. *Trends Ecol Evol.* 2005;20(6):320–7.
- Wildman D, Wu W, Goodman M, Grossman L. Episodic positive selection in ape cytochrome c oxidase subunit IV. *Mol Biol Evol.* 2002;19(10):1812–5.
- Wu C, Li W, Shen J, Scarpulla R, Limbach K, Wu R. Evolution of cytochrome c genes and pseudogenes. *J Mol Evol.* 1986;23:61–75.
- Wu S, Wu W, Zhang F, Ye J, Ni X, Sun J, Edwards S, Meng J, Organ C. Molecular and paleontological evidence for a post-cretaceous origin of rodents. *PLoS ONE.* 2012;7(10):e46445.
- Zaidi S, Hassan M, Islam A, Ahmad F. The role of key residues in structure, function, and stability of cytochrome-c. *Cell Mol Life Sci.* 2014;71:229–55.
- Zhang Z, Gerstein M. The human genome has 49 cytochrome c pseudogenes, including a relic of a primordial gene that still functions in mouse. *Gene.* 2003;312:61–72.
- Zhang Z, Gerstein M. Large-scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev.* 2004;14:328–35.
- Zhang Z, Harrison P, Liu Y, Gerstein M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* 2003;13:2541–58.
- Zhang Z, Zheng D. Pseudogene evolution in the human genome. *eLS.* 2014. doi: [10.1002/9780470015902.a0020836.pub2](https://doi.org/10.1002/9780470015902.a0020836.pub2). <http://www.els.net>.
- Zuckerandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, editors. *Evolving genes and proteins*. New York: Academic Press; 1965. p. 97–166.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
