The Molecular Taxonomy and Evolution of the Guinea Pig

W.-H. Li, W. A. Hide, A. Zharkikh, D.-P. Ma, and D. Graur

On the basis of 18 protein sequences totaling 2,413 aligned amino acid sites, it is suggested that the guinea pigs and the myomorphs (rat-like rodents) are not monophyletic. Rather, the evolutionary lineage leading to the guinea pig seems to have branched off prior to the divergence among myomorphs, lagomorphs, primates, chiropterans, artiodactyls, and carnivores. It is suggested therefore that the Caviomorpha (guinea pig-like rodents) and possibly the Hystricomorpha (porcupine-like rodents) should be elevated in taxonomic rank and conferred an ordinal status distinct from the Rodentia. This suggestion calls for a reevaluation of the morphological evolution of guinea pigs and further molecular studies on the possibility of paraphyly of the order Rodentia. If the monophyly of rodents holds, it must be concluded that the pattern of molecular evolution in many guinea pig genes has been extremely unusual and that the causes for this pattern should be sought. It is also suggested that claims of large differences in the rate of molecular evolution between guinea pigs and myomorphs may have been exaggerated in many cases as a result of an erroneous phylogenetic position for the guinea pig. The average rate of amino acid replacement in the guinea pig seems to be comparable to that in the rat and the mouse. However, the data indicate that myomorph and caviomorph genes evolve, on average, about two times faster than their human counterparts. Finally, our analysis provides evidence against the hypothesis that the gundi (an African rodent) represents the most ancient rodent lineage.

The guinea pig, *Cavia porcellus*, first caught the attention of molecular evolutionists when its insulin, a rather conservative protein, was found to have evolved much faster than other mammalian insulins, and it thus became the first and best-known exception to the molecular clock hypothesis. Indeed, guinea pig insulin differs by 18 amino acids from both human and mouse insulins, which differ from each other by only four amino acids. Moreover, unlike other mammalian insulins, guinea pig insulin can neither bind Zn^{++} ions nor form hexamers (Beintema and Campagne 1987; Watt 1985).

The guinea pig, however, would not have been the subject of this article had its insulin been the only protein that has been found to evolve at an exceptionally high rate. Rather, many other proteins of the guinea pig, such as lipoprotein lipase, glucagon, and factor IX, were also found to manifest high rates of molecular evolution. In addition, the guinea pig frequently exhibits peculiar molecular features in comparison with other eutherian mammals. For example, its gastrin is similar in length to the gastrins of marsupials but is different from the gastrins of other eutherians. As a consequence, the literature pertaining to the molecular evolution of guinea pigs abounds in references to "deviating sequences," "extremely rapid rates of evolution," and "unique evolutionary mechanisms."

In principle, when confronted with data indicating an unusually high rate of nucleotide substitution in a gene in a particular evolutionary lineage, one can invoke two types of explanation:

1. The gene is subject to relaxed functional or structural constraints in that evolutionary lineage, so that a greater proportion of mutations are selectively neutral or nearly so in comparison with homologous genes from other organisms. Such a relaxation of selective constraints can be invoked in particular cases, such as in the case of the lens crystallin of the blind mole rat, in which the loss of function has led to an increase in the rate of substitution (Hendriks et al. 1987). It is very difficult, however, to imagine a wholesale relaxa-

From the Center for Demographic and Population Genetics, University of Texas, P.O. Box 20334, Houston, TX 77225 (Li and Zharkikh), the Laboratory of Molecular Systematics, National Museum of Natural History, Smithsonian Institution, Washington, D.C. (Hide), the Department of Biochemistry and Molecular Biology, Mississippi State University (Ma), and the Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel (Graur). This paper was delivered at a symposium titled "Molecular Genetic Approaches to Phylogeny Reconstruction" sponsored by the American Genetic Association in Tucson, Arizona, April 20–21, 1991. We thank Lori Sadler for helping us in the data bank search. This study was supported by grants from NIH and the U.S.–Israel Binational Science Foundation. Address reprint requests to Dr. Li.

Journal of Heredity 1992;83:174–181; 0022-1503/92/$4.00

tion of selective pressures on as many genes as that required to explain the general tendency toward high rates of substitution in many genes of the guinea pig.

2. The rates of molecular evolution have been computed on the basis of erroneous phylogenetic assumptions. For instance, if we assume that the mouse is more closely related to man than to the rat, then we will reach the conclusion that the rate of molecular evolution in the mouse is much higher than that in the rat.

We note that estimates of the substitution rates in guinea pig proteins are usually based on the classification of the guinea pig as a rodent. What happens if this classification is incorrect? Recently, we (Graur et al. 1991) provided molecular evidence against this classification. Here we extend the data base and conduct a more detailed analysis of the molecular data. Moreover, as the guinea pig has been invariably classified as a rodent by classical taxonomists, it is important to reevaluate the morphological evidence that has been used to support this classification. However, because this is not in the domain of our expertise, we shall discuss this aspect only briefly.

Finally, we address the question of whether many guinea pig amino acid or nucleotide sequences will still exhibit a high rate of evolution if the guinea pig is reclassified outside the order Rodentia. To address this question we use marsupial or bird sequences as references to compare the substitution rates in guinea pigs, myomorphs (mice and rats), primates, and artiodactyls.

Data and Methods

To infer the taxonomic position of the guinea pig, we conducted a phylogenetic analysis of protein sequence data from caviomorphs, myomorphs, primates, artiodactyls, and some outgroup species. An outgroup species is a species for which we are certain that it had branched off before the divergence among the species under study. The outgroup may consist of more than one species. In each phylogenetic reconstruction, we chose the closest possible outgroup to the eutherians. That is, we first looked for a homologous marsupial sequence, and if no marsupial sequence was available, then we looked for a sequence from a more distant organism. Sequence data are from the GenBank, EMBL, and SwisProt protein and DNA data banks. The species used are given in Table 1.

We used protein sequences rather than DNA sequences for two reasons. First, the sequence divergence between different mammalian orders at synonymous sites is often too large, so that it is difficult to infer the number of nucleotide changes at a site. This will make phylogenetic analysis difficult. Second, in some species the DNA sequence for a protein-coding gene is not available, whereas the protein sequence is. Because DNA sequences can be translated into protein sequences, we can increase the data set by working at the protein level.

We used the maximum parsimony method (Fitch 1971) and the PROTPARS program of Felsenstein's PHYLIP package (version 3.3) to calculate (1) the number of amino acid replacements required for each of the alternative trees and (2) the number of informative sites supporting each of the trees. Informative sites are defined as those amino acid positions at which the required number of substitutions differs among the possible alternative phylogenetic trees. An informative site is said to support a tree if that tree requires the least number of substitutions at that site in comparison with the possible alternative trees. We note that at the amino acid level the number of informative sites supporting two trees may be the same, yet one of the trees may require more nucleotide substitutions. This is because a difference between two amino acids might sometimes require more than one nonsynonymous nucleotide substitution at the DNA level. We use the maximum parsimony method because it is easy to compare the results from different data sets.

Several DNA sequences are available for computing the genetic distance between guinea pigs, on the one hand, and myomorphs or humans, on the other. The number of substitutions per nonsynonymous site (K_A) and per synonymous site (K_S) were estimated by the method of Li et al. (1985).

Table 1. Species used in the study

Species	Common name	Abbreviation
Primates		
Pan troglodytes	chimpanzee	Ch
Galago crassicaudatus	galago	Ga
Homo sapiens	human	Hu
Papio anubis	olive baboon	Ob
Macaca mulatta	rhesus monkey	Rh
Ateles geoffroyi	spider monkey	Sp
Caviomorpha		
Hydrochaeris hydrochaeris	capybara	Cb
Proechimys guairae	casiragua	Cg
Chinchilla brevicaudata	chinchilla	Cl
Galea musteloides	cuis	Cu
Cavia porcellus	guinea pig	Gp
Octodon degus	octodon	Oc
Lagostomus maximus	viscacha	Vi
Myomorpha		
Mesocricetus auratus	golden hamster	Gh
Mus musculus	mouse	Ms
Rattus norvegicus	rat	Rn
Artiodactyls		
Camelus bactrianus	bactrian camel	Bc
Camelus dromedarius	camel	Cm
Bos taurus	cow	Co
Capra hircus	goat	Gt
Giraffa camelopardalis	giraffe	Gf
Sus scrofa	pig	Pg
Marsupials		
Macropus giganteus	eastern gray kangaroo	Ek
Didelphis virginiana	opossum	Op
Macropus rufus	red kangaroo	Rk
Macropus rufogriseus	red-necked wallaby	Rw
Aves		
Gallus gallus	chicken	Ck
Struthio camelus	ostrich	Os
Columba livia	pigeon	Pu
Amphibians		
Bufo japonicus	toad	Td
Insects		
Drosophila melanogaster	fruit fly	Dr

Results

Phylogenetic Analysis of Protein Sequence Data

We first consider, besides the outgroup, three eutherian groups at a time. By doing that we not only simplify the analysis but also maximize the amount of data that can be used in each analysis. In the first phylogenetic analysis we use the guinea pig (or a related caviomorph), one or more myomorphs, and one or more primates. With these three groups and an outgroup, there are three possible phylogenetic trees (Figure 1). Tree I represents the traditional taxonomical scheme, i.e., the guinea pig and the myomorphs are clustered to form a clade. In tree II, the guinea pig and the primate lineages form one clade, i.e., the two groups are evolutionarily more closely related to each other than either of them is to the myomorphs. In tree III, the guinea pig is an outgroup to both primates and myomorphs.

Table 2 presents the results of this analysis for 18 protein sequences totaling 2,413 aligned amino acid sites. There are 109 informative sites, of which 50 sites support tree III, whereas only 29 and 30 sites support tree I and tree II, respectively. The probability that tree III is wrong is roughly

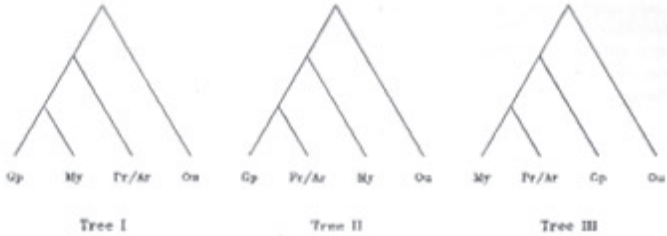

Figure 1. Three possible phylogenetic trees for the guinea pig (Gp), the myomorphs (My), the primates or the artiodactyls (Pr/Ar), and an outgroup (Ou). Tree I represents the traditional view that the caviomorphs and the myomorphs are monophyletic.

Table 2. Number of informative sites supporting each of the possible alternative phylogenetic trees for the guinea pig (Gp), primates (Pr), and myomorphs (My), and the minimum number of amino acid replacements required for each tree (in parentheses)

Protein	No. aligned sites	Tree I (Gp-My)	Tree II (Gp-Pr)	Tree III (Pr-My)	Ingroup species*	Outgroup species*
α-crystallin A chain	162	0 (33)	1 (32)	0 (33)	Gp, Hu, Rb, Ga, Ms, Ra, Gh	Op, Rk
myoglobin	153	4 (78)	2 (80)	1 (81)	Vi, Cg, Ob, Ch, Hu, Ms	Op, Rk
β-globin	145	5 (189)	0 (194)	4 (189)	Gp, Hu, Rh, Sp, Ms, Ra, Gh	Op, Rk
insulin	51	0 (31)	0 (31)	1 (30)	Gp, Hu, Ms, Ra	
"big" gastrin	31	0 (22)	0 (22)	1 (21)	Gp, Ch, Hu, Ra	Op
glucagon	29	0 (5)	0 (5)	0 (5)	Gp, Oc, Hu, Ra	Op
pancreatic polypeptide	36	0 (15)	0 (15)	0 (15)	Gp, Hu, Ms, Ra	Op
α-globin	145	4 (170)	0 (174)	5 (168)	Gp, Hu, Ob, Sp, Ms, Ra, Gh	Ek
pancreatic ribonuclease	122	1 (193)	5 (189)	1 (193)	Cl, Cu, Cb, Hu, Ms, Ra, Gh	Rk
α-lactalbumin	120	4 (148)	8 (143)	5 (147)	Gp, Hu, Ra	Rw
adrenocorticotrophin	39	0 (11)	0 (11)	1 (10)	Gp, Hu, Ms, Ra	Os
lipocortin	345	4 (179)	4 (179)	11 (172)	Gp, Hu, Ra	Pn
vasoactive intestinal peptide	28	0 (8)	0 (8)	1 (7)	Gp, Hu, Ra	Ck
lipoprotein lipase	447	2 (206)	6 (202)	12 (196)	Gp, Hu, Ms	Ck
β-nerve growth factor	121	5 (84)	1 (88)	5 (84)	Gp, Hu, Ms	Ck
myelin basic protein	101	0 (55)	0 (55)	2 (53)	Gp, Hu, Ch, Ms, Ra	Ck
vasopressin-neurophysin precursor	100	0 (60)	2 (58)	0 (60)	Gp, Hu, Ra	Td
N-*ras* protooncogene	188	0 (137)	1 (136)	0 (137)	Gp, Hu, Ms	Dr
Total	2,413	29 (1,624)	30 (1,622)	58 (1,601)		

*For species names, see Table 1.

Table 3. The effect of the outgroup on the number of informative sites supporting each of the three possible alternative trees and the minimum number of amino acid replacements required for each tree (in parentheses)

Protein	Tree I (Gp My)	Tree II (Gp Pr)	Tree III (Pr My)
Outgroup: marsupial			
α-crystallin A chain	0 (33)	1 (32)	0 (33)
β-globin	5 (189)	0 (194)	4 (189)
insulin	0 (31)	0 (31)	1 (30)
pancreatic polypeptide	0 (15)	0 (15)	0 (15)
α-globin	4 (170)	0 (174)	5 (168)
Total	9 (438)	1 (446)	10 (435)
Outgroup: chicken			
α-crystallin A chain	0 (43)	1 (42)	0 (43)
β-globin	6 (165)	2 (169)	1 (170)
insulin	1 (33)	0 (34)	1 (33)
pancreatic polypeptide	0 (15)	0 (15)	0 (15)
α-globin	1 (150)	2 (149)	5 (146)
Total	8 (406)	5 (409)	7 (407)

$P < .05$, if one assumes that each informative site has one-third probability of supporting each particular tree (Felsenstein 1985; Li and Gouy 1990). The statistical test is approximate because it assumes rate constancy and equal weights for all informative sites. Note, however, that the molecular evidence against the traditional classification (tree I) is very strong, for it is supported by less than one-third (27%) of the informative sites. Moreover, the amino acid sequence of gastrin (Shinomura et al. 1990) provides additional support for tree III. The "big" gastrins of dogs, cats, cows, sheep, goats, humans, rats, and rabbits have been shown to be 34 amino acids long. In comparison, the "big" gastrins of two New World hystricomorphs, guinea pig and chinchilla (*Chinchilla brevicaudata*), are only 33 amino acids long, lacking a glutamic acid near their NH_2-terminus, as is the gastrin of opossum (*Didelphis virginiana*), a marsupial. This makes the additional amino acid a shared-derived character of all the eutherians with the exception of the caviomorphs. Because deletions and insertions (gaps) in protein-coding regions of a gene are less frequent than are nucleotide substitutions that lead to amino acid replacement, this character should be given more weight than that given to an amino acid difference.

The outgroup species for the last eight proteins in Table 2 are from birds, toads, or *Drosophila*. This raises the question of reliability of the reference (outgroup) sequences; if the sequences under study have evolved at very different rates, then use of a distant outgroup sequence may lead to an erroneous inference on the tree topology (see Felsenstein 1978). This question can be addressed in three ways. First, for the first 10 proteins in Table 2 the outgroup species are marsupials and the numbers of sites supporting trees I, II, and III are 18, 16, and 18, respectively. Thus, for these proteins there is approximately equal support for each of the three alternative trees. Second, for each of the five proteins in Table 3 one can use either a marsupial or a chicken sequence as an outgroup. Although the number of supporting sites for each tree depends to some extent on the outgroup sequence used, trees I and III are approximately equally supported by the informative sites, regardless of whether the outgroup used is a marsupial or a bird species. Third, the reliability of an outgroup sequence depends on its distance to the ingroup sequences under study. We have therefore computed the proportion (p) of amino acid difference between an outgroup and each of the ingroup sequences (Table 4). We note that in the five cases where the outgroup is a bird species, the p values are within the range for those cases where the outgroup species is a marsupial. Because the β globin evolves at the average rate for mammalian proteins (Li and Graur 1991), we may exclude those proteins that show a p value substantially higher than that for the β globin, i.e., 30%. When these proteins

Table 4. Number and proportion (in parentheses) of amino acid differences between an outgroup (Ou) and humans (Hu), myomorphs (My), and guinea pigs (Gp)

Protein	Ou-Hu	Ou-My	Ou-Gp
Outgroup: marsupial			
α-globin	27 (0.19)	33 (0.23)	36 (0.25)
pancreatic ribonuclease	42 (0.34)	42 (0.34)	46 (0.38)
α-lactalbumin	55 (0.46)	66 (0.55)	66 (0.55)
α-crystallin	16 (0.10)	14 (0.09)	15 (0.09)
β-globin	38 (0.26)	45 (0.31)	44 (0.30)
glucagon	1 (0.03)	1 (0.03)	4 (0.14)
"big" gastrin	6 (0.19)	10 (0.32)	10 (0.32)
pancreatic polypeptide	5 (0.14)	10 (0.28)	5 (0.14)
insulin	4 (0.08)	6 (0.12)	19 (0.37)
myoglobin	22 (0.14)	35 (0.23)	25 (0.16)
Total	216	282	270
Outgroup: bird			
β-nerve growth factor	71 (0.42)	80 (0.47)	74 (0.43)
lipoprotein lipase	104 (0.23)	104 (0.23)	122 (0.27)
vasoactive intestinal polypeptide	4 (0.14)	4 (0.14)	6 (0.21)
lipocortin	98 (0.28)	109 (0.32)	98 (0.28)
myelin basic protein	33 (0.33)	35 (0.35)	32 (0.32)
Total	310	332	332

Table 5. Number of informative sites supporting each of the possible alternative phylogenetic trees for the guinea pig (Gp), artiodactyls (Ar), and myomorphs (My), and minimum number of amino acid replacements required for each tree (in parentheses)

Protein	No. aligned sites	Tree I (Gp-My)	Tree II (Gp-Ar)	Tree III (Ar-My)	Ingroup species[a]	Outgroup species[a]
α-crystallin A chain	162	1 (29)	0 (30)	0 (30)	Gp, Ms, Ra, Gh, Co, Ca, Pg	Op, Rk
myoglobin	153	4 (92)	2 (94)	0 (96)	Rk, Vi, Cg, Ms, Co, Pg	Op
β-globin	145	7 (210)	0 (217)	4 (213)	Gp, Ms, Ra, Gh, Co, Dg, Bc	Op, Rk
insulin	51	1 (31)	0 (32)	1 (31)	Gp, Ms, Ra, Bo, Pg	Op
"big" gastrin	31	0 (28)	0 (28)	1 (27)	Gp, Cl, Ra, Co, Pg	Op
glucagon	29	0 (5)	0 (5)	0 (5)	Gp, Oc, Ra, Co, Pg	Op
pancreatic polypeptide	36	0 (15)	0 (15)	0 (15)	Gp, Ms, Ra, Co, Pg	Op
α-globin	145	2 (190)	1 (191)	6 (185)	Gp, Ms, Ra, Gh, Co, Ca, Pg	Ek
pancreatic ribonuclease	122	2 (223)	3 (222)	0 (225)	Cl, Cu, Cb, Ms, Ra, Gh, Co, Ca, Gl	Rk
α-lactalbumin	120	1 (188)	6 (183)	5 (184)	Gp, Ra, Co, Dg, Ca	Rw
adrenocorticotrophin	39	0 (13)	0 (13)	1 (12)	Gp, Ms, Ra, Co, Pg	Os
lipocortin	25	1 (31)	1 (31)	0 (32)	Gp, Ra, Pg	Pm
lipoprotein lipase	447	2 (218)	9 (212)	10 (211)	Gp, Ms, Co	Ck
β-nerve growth factor	123	5 (38)	0 (43)	2 (41)	Gp, Ms, Co	Ck
vasoactive intestinal peptide	28	0 (8)	0 (8)	1 (7)	Gp, Ra, Pg	Ck
myelin basic protein	101	0 (55)	1 (54)	0 (55)	Gp, Ms, Ra, Co	Ck
vasopressin-neurophysin precursor	102	1 (58)	0 (59)	0 (59)	Gp, Ra, Co, Pg	Td
Total	1,859	27 (1,432)	23 (1,437)	31 (1,428)		

[a] For species names, see Table 1.

(i.e., pancreatic ribonuclease, α-lactalbumin, β nerve growth factor, and myelin basic protein) are excluded from the comparison, the numbers of supporting sites for the three trees become 19, 15, and 37, respectively, which still constitute strong support for tree III. It might be argued that many of the supporting sites come from lipocortin and lipoprotein lipase, but we note that these two proteins are also considerably larger than the others and that they are quite conservative proteins. In fact, their contribution to the number of informative sites supporting tree III is not significantly larger than their relative contribution to the number of aligned amino acids ($\chi^2 = 3.28$, df = 1, $P > .05$). It may be concluded therefore that the data shown in Table 2 provide strong support for the hypothesis that the guinea pig branched off earlier than the divergence occurred between primates and myomorphs.

In the second analysis, we replace the primate group by the artiodactyl group (Figure 1). The data set is now smaller (17 proteins with 1,859 aligned sites), and the number of informative sites becomes 81 (Table 5). The number of sites supporting tree III is 31, whereas the numbers of sites supporting trees I and II are 27 and 23, respectively. Although the difference in supporting sites between trees I and III is not statistically significant, this result again supports the view that the guinea pig has branched off earlier than did the myomorphs. In this case, there are two informative gaps, one supporting tree I and one supporting tree III.

In the third analysis, we consider the guinea pig, the myomorphs, the primates, and the artiodactyls together. The first three trees in Figure 2 are compatible with the traditional view that the guinea pig and the myomorphs belong to one clade. We compare these trees with tree V, which is suggested by the preceding analyses and by the study of Li et al. (1990), and also with tree IV, which is a modification of tree V with the order of myomorphs and guinea pig reversed. The minimum numbers of amino acid replacements required for each of the trees are 1,607, 1,603, 1,604, 1,607, and 1,593, for trees I, II, III, IV, and V, respectively (Table 6). Thus, tree V requires at least 10 fewer substitutions than do trees I, II, and III, which are only as parsimonious as tree IV. In this analysis, the sites used are the same as those used in the second analysis, but the differences in steps between the most-parsimonious tree and the other trees have become larger as a result of including the primate group. This result strengthens our hypothesis that the guinea pig branched off earlier than the myomorphs. These results also support the suggestion of Easteal (1990) and Li et al. (1990) that the myomorphs are an outgroup to the primates and the artiodactyls.

Finally, we examine whether or not the Ctenodactylidae, which consists of four African rodent species, belongs to the order Rodentia because this family has been suggested to represent one of the oldest rodent lineages, much older than the caviomorph lineage (Hartenberger 1985). In a recent phylogenetic analysis of α and β globin sequences, Beintema et al. (1991) suggested that the gundi (*Ctenodactylus gundi*) and the guinea pig are two early offshoots in the order of rodents, though it is not yet possible to decide whether they share a common ancestor or if they are located on separate branches. We use

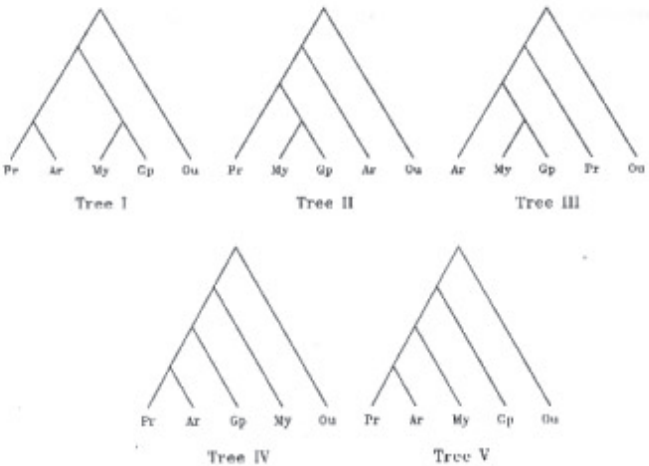

Figure 2. Five alternative phylogenetic trees out of the 15 possible ones for the primates (Pr), the artiodactyls (Ar), the myomorpha (My), the guinea pig (Gp), and an outgroup (Ou). Trees I, II, and III are compatible with rodent monophyly.

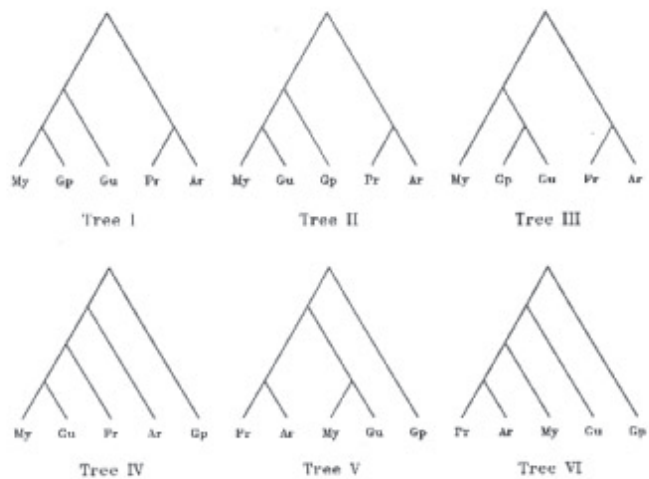

Figure 3. Six alternative rooted trees out of 105 possible ones for the gundi (Gu), the guinea pig (Gp), the myomorphs (My), the primates (Pr), and the artiodactyls (Ar). Trees I, II, and III are compatible with rodent monophyly. In the analysis, kangaroos are used as an outgroup.

α and β globins, myoglobin, and αA-crystallin; the total number of aligned amino acid sites is 601. Only in the following eutherian mammals have all these proteins been sequenced: gundi, human (representing the primates), guinea pig, mouse (representing the myomorphs), and bovine and pig (representing the artiodactyls). (Actually, the guinea pig myoglobin has not been sequenced yet, so we use instead the myoglobin from the closely related casiragua, *Proechymys guairae*.) As an outgroup we use red kangaroo myoglobin, β globin, and αA-crystallin, and eastern grey kangaroo α globin. Our analysis shows that tree VI in Figure 3 is the most parsimonious, requiring at least 375 amino acid replacements. Trees IV and V are the second most parsimonious, each requiring at least 376 replacements. Trees I, II, and III, which represent the three possible arrangements under the assumption of rodent monophyly, require at least 378, 379, and 377 replacements, respectively. Although the differences in steps among these alternative trees are too small to be significant, the results provide further evidence against rodent monophyly and suggest that, like the guinea pig, the gundi might also have branched off earlier than the divergence among the myomorphs, the primates, and the artiodactyls. Note that there is no evidence for Hartenberger's (1985) suggestion that the gundi lineage is much older than the guinea pig lineage.

Molecular Clocks

Table 4 shows the number of amino acid differences between an outgroup, on the one hand, and humans, myomorphs, or guinea pigs, on the other. We divide the data into two groups. In the first the outgroup is a marsupial, and in the second the outgroup is a bird. There is no significant difference in the total number of replacements between the myomorph lineage and the guinea pig lineage. In several proteins, e.g., insulin and lipoprotein lipase, guinea pigs have accumulated an excess of amino acid replacements in comparison to the myomorphs, but in some other cases, e.g., lipocortin and β nerve growth factor, the opposite is observed. Thus, the rate of amino acid replacement is approximately the same in guinea pigs and myomorphs. In 10 out of 15 comparisons, the rate of amino acid replacement is higher in myomorphs than in humans. In 11 out of 15 comparisons, the rate is higher in guinea pigs than in humans. Therefore, the rate of amino acid replacement is higher in both the guinea pig and myomorph lineages than in the human lineage.

Table 7 shows the number of synonymous (K_S) and nonsynonymous (K_A) substitutions between guinea pigs, humans, and myomorphs. In almost all cases, the K_A and K_S values between guinea pigs and humans are smaller than the corresponding values between guinea pigs and myomorphs. The only exceptions are the K_S values for pancreatic polypeptide and insulin-like growth factor I. The weighted K_A and K_S values in this table are almost the same as those in Graur et al. (1991), although three new genes (lipocortin, transglutaminase, and islet amyloid protein) with a total of 3,393 nucleotides were added to the comparison. The length of each

Table 6. Minimum number of amino acid replacements required for each of the phylogenetic trees in Figure 2

Protein	Tree I	II	III	IV	V
lipoprotein lipase	237	238	235	231	226
α-crystallin A chain	40	40	40	40	41
α-globin	216	216	217	220	214
β-globin	243	242	239	245	240
α-lactalbumin	205	208	208	202	204
vasoactive intestinal peptide	8	8	8	8	7
β-nerve growth factor	42	41	41	45	43
insulin	28	28	28	29	28
"big" gastrin	31	31	30	30	29
adrenocorticotropin	13	13	13	13	12
pancreatic ribonuclease	259	255	261	258	261
pancreatic polypeptide	17	17	17	17	17
lipocortin	35	34	34	34	34
myoglobin	105	105	105	107	109
vasopressin-neurophysin precursor	64	64	64	64	64
glucagon	5	5	5	5	5
myelin basic protein	59	58	59	59	59
Total	1,607	1,603	1,604	1,607	1,593

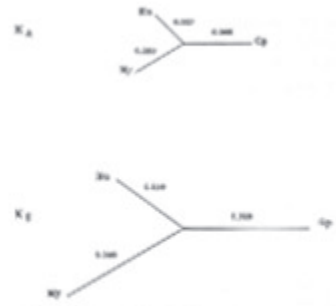

Figure 4. Branch lengths from a common node to humans (Hu), myomorphs (My), and guinea pigs (Gp) computed from the number of nucleotide substitutions per nonsynonymous site (K_A) and per synonymous site (K_S) given in Table 7.

Table 7. Number of substitutions per nonsynonymous site (K_A) and synonymous site (K_S) between guinea pigs and humans (Gp-Hu), guinea pigs and myomorphs (Gp-My), and humans and myomorphs (Hu-My)

Gene	N[a]	K_A Gp-Hu	Gp-My	Hu-My	K_S Gp-Hu	Gp-My	Hu-My
preproinsulin	330	0.208	0.211	0.102	0.700	0.854	0.673
apolipoprotein E	906	0.149	0.176	0.175	0.387	0.771	0.618
preproglucagon	540	0.044	0.061	0.043	0.484	0.660	0.560
α-lactalbumin	411	0.170	0.241	0.174	0.550	1.112	0.848
lipoprotein lipase	1,535	0.065	0.071	0.029	0.653	0.671	0.607
pancreatic polypeptide	213	0.149	0.193	0.155	0.958	0.789	0.486
N-ras protooncogene	564	0.002	0.009	0.002	0.284	0.516	0.462
factor IX	804	0.116	0.135	0.097	0.366	0.676	0.525
insulin-like growth factor I	604	0.210	0.292	0.224	0.861	0.946	0.666
lipocortin	1,038	0.050	0.073	0.056	0.817	0.764	0.588
transglutaminase	2,029	0.093	0.105	0.096	0.506	0.649	0.627
islet amyloid protein[b]	276	0.184	0.251	0.188	0.563	0.837	0.816
Weighted average		0.103	0.126	0.097	0.545	0.723	0.613

[a] N = number of nucleotides compared.
[b] The caviomorph sequence is from *Octodon degus*.

branch connecting the common node and each taxa can be calculated by the method of Fitch and Margoliash (1967) (Figure 4). Under the traditional view that the guinea pig and the myomorphs belong to one clade with the root of the tree on the branch leading to humans, the rate of nucleotide substitution between the guinea pig and the myomorph lineages would be about two times faster than that in the human lineage. Under the new proposed phylogeny with the root on the branch leading to the guinea pig, the rate in the myomorph lineage would be about two times faster than that in the human lineage. Under this phylogeny, whether the rate of nucleotide substitution is higher in the guinea pig lineage than in the human lineage depends on the location of the root, but, as noted above (Table 4), the rate of amino acid replacement is higher in the guinea pig lineage than in the human lineage. Therefore, we may conclude that the molecular clock runs faster in the guinea pig and myomorph lineages than in the human lineages and that the rate-constancy assumption does not hold for all these three lineages, although it may hold for the guinea pig and myomorph lineages.

Discussion

Mammalian Classification

The order Rodentia is traditionally divided into three main extant suborders: the Sciuromorpha (squirrel-like rodents), the Myomorpha (rat-like rodents), and the Hystricomorpha (porcupine-like rodents). While each of the first two suborders is generally considered to be monophyletic, the New World families of the Hystricomorpha, i.e., the Caviomorpha (guinea pig-like rodents), are sometimes thought to have evolved independently of the Old World hystricomorph rodents, and the caviomorphs are sometimes conferred an independent subordinal status (Nowak and Paradiso 1983; Romer 1966). Today, many rodent taxonomists recognize only two suborders, the Hystricognathi, which includes the hystricomorphs and the caviomorphs, and the Sciurognathi, which includes the myomorphs and the sciuromorphs.

Our molecular analyses suggest that the guinea pig does not belong to the same order as the myomorphs but represents a separate evolutionary lineage that diverged before the separation of the myomorph rodents from the primates and the artiodactyls. It is not yet possible to decide which species besides the caviomorphs belong to the new mammalian order represented by the guinea pig. Though controversial, the caviomorphs are commonly regarded as monophyletic with the Old World hystricomorphs, and this claim is supported by limited molecular data (Fitch and Beintema 1990; Sarich 1985). We therefore tentatively suggest that the hystricomorphs represent a new order distinct from the Rodentia. If, as suggested by recent data (Easteal 1990; Li et al. 1990; Mindell D, personal communication), the myomorphs branched off before the carnivores, artiodactyls, lagomorphs, chiropterans, and primates diverged, then the new order would represent one of the earliest divergence events in the evolutionary history of the eutherian mammals.

Morphology in Rodent Classification

Because the guinea pig has been invariably classified as part of the order Roden-

tia, our results point toward the need to reevaluate the morphological basis of rodent taxonomy. We emphasize that we are not professionally proficient in mammalian morphology and anatomy, and thus in the following we shall mostly raise questions rather than provide solutions.

The crucial characters used to distinguish rodents from other mammalian orders are those related to gnawing (Hartenberger 1985). Other characters, such as the separation of the optic foramina and the size of the orbitosphenoid, may characterize the majority of rodents, but no character unrelated to the habit of gnawing can be shown to be present in all the rodents and absent in all other mammalian orders. In fact, most students of rodent taxonomy regard the front incisors that have developed into an effective chisel-like gnawing apparatus as the "key character of the group" (Romer and Parsons 1977). This is so much so that Ellerman (1966) stated not only that "a rodent becomes a rodent because it gnaws" but also that "if *Rattus*, for example, had never taken to gnawing, it would not be classed in the present Order today." In particular, the following three dental characters are thought to be autapomorphic (uniquely derived) and are used to support the monophyly of the group (Hartenberger 1985):

1. One upper and one lower pair of incisors, enlarged and evergrowing;
2. Absence of canines and anterior premolars creating a long diastema between the incisors and the posterior premolars on both jaws;
3. Enamel of the incisors restricted to the buccal (anterior) face.

None of these features, however, is exclusively associated with rodents. For example, the same dental features can be found in the aye-ayes *Daubentonia* (a primate) and in wombats (marsupials). As far as the long diastema is concerned, the same character is found in the fossil reptiles: *Bienotherium*, *Likhoelia*, and *Oligokypus* (Hartenberger 1985). Moreover, the dental formula in rodents is quite variable—in some cases changing throughout ontogeny—and thus cannot be used in a cladistic analysis to establish rodent monophyly. Finally, when the Lagomorpha was removed from the Rodentia, one of the characters used to establish the monophyly of rodents to the exclusion of the rabbits was the mechanism of chewing (exclusively lateral mastication in the lagomorphs, lateral and anterio-posterior movements of the mandible in the rodents). We note, however, that some mammals other than rodents, such as elephants and hyraxes, masticate their food as rodents do.

It also seems to us difficult to argue too strongly for the monophyly of the Rodentia when there are tremendous difficulties in establishing monophyly of the suborders within the Rodentia (Hartenberger 1985). Indeed, it has been suggested that the hystricomorph condition evolved independently no less than eight times, and the sciuromorph condition six times (Wood 1974; Young 1981). Such a prevalence of parallelism may render many morphological characters useless for the purpose of taxonomy. We must therefore conclude that, on the basis of morphological characters, the monophyly of the Rodentia has not been established beyond doubt.

Unfortunately, we fared no better in finding morphological characters that may be used to support our proposed molecular phylogeny. We found only two such characters, neither very convincing:

1. Presence of enamel tubules in hystricomorphs, marsupials, and reptiles and absence of such tubules in myomorphs, sciuromorphs, and all other eutherian orders including extinct ones (Sahni 1985).
2. The scale pattern of the hystricomorph hair is wavy or streaked, similar to the condition found in the opossum and the platypus. In contrast, the scale pattern of most other eutherians (including myomorphs and sciuromorphs) is diamond-petal or transitional-petal (Brunner and Coman 1974).

Differential Rates of Molecular Evolution

Our proposal for a new taxonomic assignment for the guinea pig resolves or ameliorates many of the paradoxes associated with the evolution of its genes. For example, consider the case of lipocortin. Under the assumption that the guinea pig is a rodent or at least a sister group of the rodents, one must assume that at least 15 amino acid replacements have occurred in parallel in pigeons and in guinea pigs. The number of parallel substitutions is reduced to eight if the new taxonomic position of the guinea pig is accepted.

In many studies the rate of substitution in guinea pig genes relative to that in myomorph species has been computed by using the relative rate test (Sarich and Wilson 1973) with human or bovine genes as an outgroup; that is, tree 1 in Figure 1 was assumed to represent the true phylogeny. For many genes the conclusion was that the guinea pig lineage evolves much faster than the myomorph lineage. For example, the α and β globin chains were found to have evolved approximately two to four times faster in the guinea pig than in mice or rats (Shoshani et al. 1985). In contrast, under our proposed phylogenetic position for the guinea pig, the two sequences seem to have evolved at about the same rate in both the caviomorph and the myomorph lineages.

We therefore conclude that claims of large differences in the rate of evolution between guinea pigs and myomorphs may have been exaggerated in many cases as a result of an erroneous phylogenetic assumption. However, there is evidence that both the caviomorph and myomorph lineages have evolved faster than the human lineage (Tables 4 and 7). Note, however, regardless of the evolutionary position of the guinea pig, some guinea pig proteins, e.g., insulin and lipoprotein lipase, have clearly evolved at exceptionally high rates.

In conclusion, the data we have compiled suggest either that the guinea pig is not a rodent or that many guinea pig genes have evolved at exceptionally high rates and that their pattern of molecular evolution is highly unusual. In the latter case, the causes for such unusual patterns should be sought. The fact is that many guinea pig genes are very divergent from their myomorph orthologs. Therefore, even if the caviomorphs and the myomorphs do turn out to be monophyletic, the parsimony analyses (Tables 2, 5, and 6) suggest that the caviomorph, the myomorph, and the primate lineages are close to a trichotomy, and the genetic distances between caviomorphs and myomorphs (Table 7) are large enough to warrant a separate ordinal status for the Caviomorpha.

References

Beintema JJ and Campagne RN, 1987. Molecular evolution of rodent insulins. Mol Biol Evol 4:10–18.

Beintema JJ, Rodewald K, Brammizer G, Czehunnick J, and Goodman M, 1991. Studies on the phylogenetic position of the Ctenodactylidae (Rodentia). Mol Biol Evol 8:151–154.

Brunner H and Coman BJ, 1974. The identification of mammalian hair. Melbourne: Inkata Press.

Easteal S, 1990. The pattern of mammalian evolution and the relative rate of molecular evolution. Genetics 124:165–173.

Ellerman JR, 1966. The families and genera of living rodents. Codicote, England: Wheldon and Wesley.

Felsenstein J, 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410.

Felsenstein J, 1985. Confidence limits on phylogenies with a molecular clock. Syst Zool 34:152–161.

Fitch WM, 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst Zool 20:406–416.

Fitch WM and Beintema JJ, 1990. Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense branching as exemplified by ribonuclease. Mol Biol Evol 7:438–443.

Fitch WM and Margoliash E, 1967. Construction of phylogenetic trees. A method based on mutation distances as estimated from cytochrome *c* sequences is of general applicability. Science 155:279–284.

Graur D, Hide WA, and Li W-H, 1991. Is the guinea-pig a rodent? Nature 351:649–652.

Hartenberger J-L, 1985. The order Rodentia: major questions on their evolutionary origin, relationships and suprafamilial systematics. In: Evolutionary relationships among rodents: a multidisciplinary analysis (Luckett WP and Hartenberger J-L, eds). New York: Plenum Press; 1–33.

Hendriks W, Leunissen J, Nevo E, Bloemendal H, and de Jong W, 1987. The lens protein αA-crystallin of the blind mole rat, *Spalax ehrenbergi*: evolutionary change and functional constraints. Proc Natl Acad Sci USA 84: 5320–5324.

Li W-H and Gouy M, 1990. Statistical tests of molecular phylogenies. In: Methods in enzymology. Molecular evolution: computer analysis of protein and nucleic acids sequences (Doolittle RF, ed). New York: Academic Press; 645–659.

Li W-H, Gouy M, Sharp PM, O'hUigin C, and Yang Y-W, 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. Proc Natl Acad Sci USA 87:6703–6707.

Li W-H and Graur D, 1991. Fundamentals of molecular evolution. Sunderland, Massachusetts: Sinauer.

Li W-H, Wu C-I, and Luo C-C, 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174.

Nowak RM and Paradiso JL, 1983. Walker's mammals of the world. Baltimore and London: Johns Hopkins University Press.

Romer AS, 1966. Vertebrate paleontology. Chicago: University of Chicago Press.

Romer AS and Parsons TS, 1977. The vertebrate body. Philadelphia: Saunders.

Sahni A, 1985. Enamel structure of early mammals and its role in evaluating relationships among rodents. In: Evolutionary relationships among rodents: a multidisciplinary analysis (Luckett WP and Hartenberger J-L, eds). New York: Plenum Press; 133–150.

Sarich VM, 1985. Rodent macromolecular systematics. In: Evolutionary relationships among rodents: a multidisciplinary analysis (Luckett WP and Hartenberger J-L, eds). New York: Plenum Press; 423–452.

Sarich VM and Wilson AC, 1973. Generation time and genomic evolution in primates. Science 179:1144–1147.

Shinomura Y, Eng J, Rattan SC, and Yalow RS, 1990. Opossum (*Didelphis virginiana*) "little" and "big" gastrins. Comp Biochem Physiol 96B:239–242.

Shoshani J, Goodman M, Czelusniak J, and Braunitzer G, 1985. A phylogeny of Rodentia and other eutherian orders: parsimony analysis utilizing amino acid sequences of alpha and beta hemoglobin chains. In: Evolutionary relationships among rodents: a multidisciplinary analysis (Luckett WP and Hartenberger J-L, eds). New York: Plenum Press; 191–210.

Watt VM, 1985. Sequence and evolution of guinea pig preproinsulin DNA. J Biol Chem 260:10926–10929.

Wood AE, 1974. The evolution of the Old World and New World hystricomorphs. Symp Zool Soc London 34:21–60.

Young JZ, 1981. The life of vertebrates. Oxford, England: Clarendon Press.