Complete Mitochondrial DNA Sequence of the Fat Dormouse, *Glis glis*: Further Evidence of Rodent Paraphyly

Aurelio Reyes,* Graziano Pesole,† and Cecilia Saccone‡

*Centro di Studio sui Mitochondri e Metabolismo Energetico, Bari, Italy; †Department of Biology, University of Basilicata, Potenza, Italy; and ‡Department of Biochemistry and Molecular Biology, University of Bari, Bari, Italy

The complete mitochondrial genome of the fat dormouse, *Glis glis*, has been sequenced (16,602 bp). A total of 23 complete mitochondrial mammalian genomes have been taken into account for phylogenetic reconstruction. Phylogenetic analyses were performed with parsimony, distance (stationary Markov model), and maximum-likelihood methods. In all cases, data strongly support the paraphyly of rodents, with dormouse and guinea pig in a different clade from rat and mouse, reaching bootstrap values of 95%. Rodent monophyly and the existence of Glires (Rodentia and Lagomorpha) are weakly supported, with maximum bootstrap values of 11% and 8.6%, respectively. This result agrees with the analyses of isochore patterns in the nuclear genome and the B2 and B2-like retroposons, which show a close relationship between dormice and guinea pigs rather than between dormice and rats and mice.

Introduction

Rodents appeared on earth about 55 MYA according to paleontological data (Hartenberger 1985, 1996). Soon afterward, an explosive radiation of rodents took place, with an average of 2.8 families and 7.3 genera per Myr (Hartenberger 1996). Thus, nowadays, the order Rodentia includes almost half of the mammals’ 1.800–2,300 species, grouped in 30–33 extant families (Wilson and Reeder 1993), which colonize almost all terrestrial environments: mountains, rivers, deserts, ice lands, and cities (Carleton 1984; Hartenberger 1985).

Based on morphological features, several suprafamilial classifications have been established. According to Tullberg (1899), rodents are classified into two groups, Sciurognathi and Hystricognathi, based mainly on the sciurognathy or hystrycognathy of the jaw and developmental features. The former group was split into two infraorders, Sciuromorpha and Myomorpha. The infraorder Myomorpha includes almost 80% of all rodents, grouped in different superfamilies, such as Muridea, Dipodoidea, Geomyoidea, and Gliridea.

Despite the great diversity in morphology, behavior, and ecology, this order has been unanimously accepted as monophyletic since its first classification in the early 19th century on account of their cranial, dental, postcranial, and soft anatomical attributes (Hartenberger 1985; Luckett and Hartenberger 1993). Nevertheless, some of the characters used in morphological classification of rodents are supposed to be originated by convergent or parallel evolution. In other cases, the determination of the polarity of evolution is not possible for many characters, as morphological modifications in both directions with regard to time are very frequent in rodents (Hartenberger 1985; Li et al. 1992).

Rodent monophyly was challenged for the first time by the molecular approach at the beginning of this decade. Graur, Hide, and Li (1991) suggested, based on the analysis of 15 nuclear protein sequences, that genetic distances between the Cavioida guinea pig and the Muroida rat and mouse were large enough to postulate rodent paraphyly. Later on, other molecular surveys of both nuclear and mitochondrial genes were carried out in order to test the monophyly or paraphyly of this order. However, in most of those surveys, only three rodent species were analyzed: rat and mouse, representing Muroida, and guinea pig, representing Cavioida. Monophyly of rodents was supported in some surveys (Hasegawa et al. 1992; Martignetti and Brosius 1993; Cao et al. 1994; Kuma and Miyata 1994; Frye and Hedges 1995; Porter, Goodman, and Stanhope 1996), while in others, molecular evidence contradicted the traditional classification of rodents and suggested rodent paraphyly (Graur et al. 1992; Li et al. 1992; Ma et al. 1993; Wolf et al. 1993; Noguchi et al. 1994; D’Erchia et al. 1996; Janke, Xu, and Arnason 1997).

The family Gliridae represents a very interesting case of relationships among rodents. According to paleontological data, it could be considered one of the most primitive rodents, along with Ctenodactylidae and Geomyidae. Fossil records of Gliridae are dated in the middle Eocene (45 MYA), while other rodent ancestors were dated in the Eocene–Oligocene boundary, 35 MYA (Hartenberger 1985). Furthermore, the oldest known glirid fossil, *Eogliravus* sp., has an enlarged infraorbital foramen as in the hystricomorphy, but in more advanced genera, the lateral portion of the masseter has extended anteriorly to produce a pseudomyomorph pattern, also found in modern dormice. The existence of this pseudomyomorph pattern of Gliridae led taxonomists to include this family within the group Myomorpha (Carrol 1988; Wahlert and Savitske 1993; Vianey-Liaud 1994).

Molecular data available on Gliridae species are scarce. Determination of phylogenetic relationships between this family and other rodents has been attempted by means of the analysis of the partial mitochondrial 12S rRNA sequences (Catzeﬁlis et al. 1995) and isochore patterns in the nuclear genome (Sabeur et al. 1993) and the B2 and B2-like retroposons (Serdobova and Kramarov 1993). Mitochondrial rRNA analysis did not clarify the branching pattern of major lineages of rodents. Isochore analysis detected similarities between Gliridae...
and both Muridae and Caviidae species, while retroposon analysis detected a close relationship only between Gliridae and Caviidae.

Mitochondrial DNA has proven to be a powerful tool for phylogenetic reconstruction, especially when complete genomes are considered. Another advantage is that mtDNA genes are actually orthologous. In the present study, we examine the phylogenetic relationships among four rodent species in the context of mammalian evolution by adding the complete mitochondrial sequence of the fat dormouse *Glis glis*. The addition of this species, which was initially considered a myomorph and at the same time has some similarities to hystricomorph species, could shed light on the controversy of rodent monophyly or paraphyly. In addition, a superorder Glires (Rodentia and Lagomorpha) has also been revised.

Material and Methods

Enriched mtDNA was isolated from 4 g of frozen tissues (kidney, liver, spleen, and heart) of *G. glis*, wild-caught near Montpellier, France (specimen number V-779 of E. Catzeflis), according to previously described methods for mammalian species (Arnason, Gullberg, and Widegren 1991). Restriction fragments generated by single or double digestions with *Bam HI, Eco RI, Hind III, Psi I*, and *Xba I* were ligated into M13mp19 and cloned in *Escherichia coli* XL1-Blue. Single-stranded DNA from these fragments, containing the whole mitochondrial genome, were sequenced following the di-deoxy termination technique (Sanger 1981) with [α-35S]dATP using both universal and numerous specific oligonucleotide primers.

The mtDNA sequence of *G. glis* has been deposited in the EMBL database with accession number AJ001562. The following complete mammalian mitochondrial DNA sequences available in databases (EMBL release 49) were taken into account in the evolutionary analyses: human, V00662 (Anderson et al. 1981); common chimpanzee, D38116 (Horai et al. 1995); pigmy chimpanzee, D38113 (Horai et al. 1995); gorilla, D38114 (Xu and Arnason 1996); orang-utan, D38115 (Horai et al. 1995); gibbon, X99256 (Arnason, Gullberg, and Xu 1996); horse, X79547 (Xu and Arnason 1994); donkey, X97337 (Xu, Gullberg, and Arnason 1996); Indian rhinoceros, X97336 (Xu, Janke, and Arnason 1996); harbor seal, X63726 (Arnason and Johnsson 1992); gray seal, X72004 (Arnason et al. 1993); cat, U20753 (Lopez, Cevario, and O’Brien 1996); fin whale, X61145 (Arnason, Gullberg, and Widegren 1991); blue whale, X72204 (Arnason and Gullberg 1993); cow, V00654 (Anderson et al. 1982); rabbit and guinea pig (D’Erchia et al. 1996); rat, X14848 (Gadaleta et al. 1989); mouse, V00711 (Bibb et al. 1981); hedgehog, X88898 (Krettek, Gullberg, and Arnason 1995); opossum, Z29573 (Janke et al. 1994); and platypus, X83427 (Janke et al. 1996). Noneutherian sequences were always used as mammalian outgroups. Chicken (X52392) (Desjardins and Morais 1990) and frog (X02890) (Roe et al. 1985) sequences were included as nonmammalian species in order to be used as further outgroups.

Concatenated supergenes for ribosomal 12S and 16S genes and protein-coding genes codified in the H-strand were multialigned using the PILEUP program, and manual adjustments were made when necessary by means of the LINEUP program (Devereux, Haeberli, and Smithies 1984). In the case of protein-coding genes, nucleotide alignment was guided by the corresponding amino acid alignment. Based on these multialignments, different phylogenetic analyses, both deterministic and stochastic, were performed. In all cases, gaps were excluded from the phylogenetic analyses. Maximum-parsimony (MP) analysis was conducted on amino acid sequences using the PAUP program (Swofford 1993). Genetic distances were calculated using the Markov method (Saccone et al. 1990) on the ribosomal supergene and synonymous codon positions of the protein-coding supergene. Synonymous codon positions have not been considered because of the lack of stationarity (biased base composition) and their fast substitution rates, which lead to a nucleotide divergence above the saturation level for many interspecies comparisons. These distances were used for phylogenetic reconstruction using the neighbor-joining (NJ) method (Saitou and Nei 1987). Also, maximum-likelihood (ML) approaches were attempted on ribosomal and amino acid multialignments using the DNAML and PROTML programs from the PHYLIP (Felsenstein 1991) and MOLPHY (Adachi and Hasegawa 1995a) packages, respectively.

Results

The mtDNA of the fat dormouse is 16,602 bp long and encodes for 13 protein-coding genes, 2 ribosomal rRNAs, and 22 tRNAs. The control region is 1,157 nt long and does not contain repetitive motifs. The organization of this genome conforms to that of other mammalian species. The base composition of the L-strand is A, 30.6%; C, 24.6%; G, 11.8%; T, 33.0% for the H-strand and coding genes and A, 35.6%; C, 21.0%; G, 17.0%; T, 24.0% for the L-strand. These values are near the averages of other mammalian mtDNAs.

The phylogenetic tree based on concatenated mitochondrial amino acid sequences obtained by MP analysis is shown in figure 1. A clade comprising Primates, Perissodactyla, Carnivora, Cetacea, Artiodactyla, Lago- morpha, and the rodents guinea pig and dormouse is supported by a bootstrap value of 63%. The close relationships between Perissodactyla and Carnivora, Cetacea and Artiodactyla, and these two superordinal groups are supported with high bootstrap values. However, the position of Lagomorpha and the two rodents, dormouse and guinea pig, remains unsettled, as a polytomy among dormouse, guinea pig, Lagomorpha, Primates, and the clade containing Perissodactyla, Carnivora, Cetacea, and Artiodactyla is found. The muroid rodents, rat and mouse, are separated from the two other rodents, with a bootstrap value of 63%. Monophyly of rodents would be supported by a bootstrap value of 4.5%, while the relationship between dormouse and the
other muroids is supported by a bootstrap value of 14%. The existence of Glires is supported by a value of only 8.7%. Hedgehog is placed at the base of the eutherian tree, with a bootstrap value of 100%.

The same amino acid sequence data set was also used to build a phylogenetic tree by means of ML using PROTML (fig. 2). Two rodent clades were observed, one containing rat and mouse and the other containing dormouse and guinea pig. Thus, rodent paraphyly is supported with a high bootstrap value (96%), while neither rodent monophyly nor dormouse clustering with rat and mouse and separated from guinea pig was supported by the data. Rabbit appears in a cluster with dormouse and guinea pig that is supported with a bootstrap value of 66%. No changes are detected regarding the relationships among the other mammalian species.

Concatenated nucleotide sequences of protein-coding genes were used to build a tree (fig. 3) using the neighbor-joining method on genetic distances calculated on synonymous codon positions with the Markov model. This method allows a better resolution of the clades, as no polytomies are observed and bootstrap values are rather high. Rat and mouse branched earlier than dormouse and guinea pig. Rodent paraphyly based on the existence of two different clades is supported with a high bootstrap value (80%). Rodent monophyly is supported by a bootstrap value of only 11%, while the dormouse clustering with rat and mouse and separated from guinea pig has a bootstrap support of 7%. Glires were not supported by the data. Rabbit appears in a cluster with ferungulates (cetaceans, artiodactyls, perissodactyls, carnivores). The relationships among the other species of mammals described above were confirmed.

The phylogenetic tree obtained by applying the stationary Markov model and NJ on the rRNA concatenated supergene is shown in figure 4. Again, two rodent clades were observed, one of them containing rat and mouse and the other containing dormouse and guinea pig (bootstrap value 82%), giving further support to the results described above. The alternative rodent monophyly is supported by a bootstrap value of only 13%, and the clustering together of rat, mouse, and dormouse is supported by a value of 2%. Glires are supported by a 3% value. In this case, rabbit branched off after dormouse and guinea pig. Relationships among mammalian species remain invariant and in accord with results obtained with other methods.

The same analysis was also carried out using frog and chicken in order to include more distant outgroups,
since rodents have been found to be close to the base of the tree. In this case, the topology of the tree has not changed, and the existence of two clades in rodents is also supported with high bootstrap values (73% for the protein-coding supergene and 71% for the rRNA supergene).

Due to the fact that hedgehog shows an atypical mammalian sequence, analyses were also carried out without this species. When this species was not taken into account, tree topology remained constant but bootstrap values were higher. Thus, rodent paraphyly is supported by a bootstrap value of 94%, while rodent monophyly is supported only by 2% based on the protein-coding supergene. The resulting values for the rRNA supergene are 83% and 7% for rodent paraphyly and monophyly, respectively. Glires had no support from these data.

Maximum-likelihood analysis on rRNA or protein sequences gave results that supported the division of rodents into the two groups previously described. The relationships between the other mammalian lineages did not change.

Discussion

Traditionally, the determination of phylogenetic relationships among mammalian species has been attempted using morphological data; however, advances in molecular biology have made it possible to incorporate molecular data into mammalian classifications. Nevertheless, rather than converging in their results, morphological and molecular data have sometimes shown conflicting viewpoints on the monophyly or paraphyly of different orders and/or superorders. The monophyly of the order Rodentia and its clustering with Lagomorpha in the cohort Glires has been one of the most widely discussed points in mammalian phylogeny.

In the present study, we analyzed the relationship among four rodent species in the context of mammalian phylogeny using complete mitochondrial genomes. Based on three different methodological approaches (figs. 1–4), we have always obtained the result that fat dormouse clusters with guinea pig apart from rat and mouse. Bootstrap values supporting rodent paraphyly are high (82%–96%) and very different from those supporting monophyly (2%–11%).

These results are clearly against the traditional morphological classification, where rodent monophyly and the close relationship between murids (rat and mouse) and glirids (dormouse) have been suggested, based mainly on the masticatory apparatus and cranial and developmental features (Hartenberger 1985; Luckett and
Muridae and other families such as Caviidae or Sciuridae, rodent species showed important differences between asymmetry (mean minus modal value) with the muroids. However, dormouse genome shares a low buoyant density of 1.6991; hence, it is closer to the one obtained for guinea pig (1.6989) than to those described for rat and mouse (1.7008 and 1.7006, respectively). However, dormouse genome shares a low asymmetry (mean minus modal value) with the muroids. Analyses on isochore patterns carried out on different rodent species showed important differences between Muridae and other families such as Caviidae or Sciuridae, but Gliridae were not considered. On the other hand, B2 and B2-like retroposons have been shown to be present in Dipodidae and Muridae but not in Gliridae and Caviidae, which would support the results obtained here with complete mitochondrial genomes. This contrasts with data from retroposon BC1, present both in murids and guinea pig, thus demonstrating the unreliability of phylogenetic information based on presence/absence of mobile elements.

Another controversial point is the existence of the cohort Glires. According to morphological data, Rodentia and Lagomorpha constitute a monophyletic group with well-defined characters (Hartenberger 1996). Nevertheless, in accord with previous results (D’Erchia et al. 1996; Janke, Xu, and Arnason 1997), we have not obtained support for this hypothesis based on mitochondrial complete genomes (maximum bootstrap value 8.6%). Nuclear data (Graur, Duret, and Gouy 1996) also contradict the traditional classification of Glires. Nevertheless, the phylogenetic position of rabbit is not yet completely defined, and thus we obtain different results when using protein-coding genes or ribosomal genes (figs. 2–4).

It is remarkable that all the analyses carried out so far on complete mitochondrial genomes, even using different species-sampling and analytical models, all find the hypothesis of rodent polyphyly more likely (D’Erchia et al. 1996; Cao, Okada, and Hasegawa 1997; Janke, Xu, and Arnason 1997). However, some authors claim that the results are still nonsignificant enough to question the established rodent monophyly. In particular, Sullivan and Swoford (1997), even when adopting an heterogeneous rate model accounting for a different variability of nucleotide sites along the sequence, again find the rodent polyphyly tree most likely although not significantly “better” than the best rodent monophyly tree. Indeed, the use of more sophisticated models, such as that in Sullivan and Swoford (1997), has two main drawbacks: (1) If the assumed model describing site rate variability does not fit real data, the results can be even less significant than those obtained using an oversimplified model assuming equal rates among sites. This was previously clearly demonstrated in Pesole et al. (1995). (2) Since the size of statistical fluctuations is proportional to the number of parameters used in the analytical model, if the statistical fluctuations are too high, no significant phylogeny can be obtained from the molecular data. Indeed, this happens in the analyses reported by Cao, Okada, and Hasegawa (1997) and Sullivan and Swoford (1997) when, even using a large data set such as the complete mitochondrial genome, no significant phylogeny, either supporting or refuting rodent monophyly or polyphyly, can be drawn.

We would like to stress that regarding the other eutherian orders, the results are in agreement with previous surveys based on mitochondrial genomes (D’Erchia et al. 1996; Janke, Xu, and Arnason 1997). Moreover, the addition of new sequences has helped to settle the previous uncertain position of perissodactils (D’Erchia et al. 1996), now clustered with Carnivora with high bootstrap values, in agreement with previous findings (Xu, Janke, and Arnason 1996).
Based on the data available so far, we can conclude that within the order Rodentia there are two well-defined groups which are supported by high bootstrap values obtained with different methodological approaches. One of these groups contains murids (rat and mouse), while the other includes dormouse and guinea pig. Nevertheless, some authors have claimed that taxonomic sampling might have a major impact on phylogenetic inferences (Lecointre et al. 1993; Philippe and Douzery 1994; Adachi and Hasegawa 1995b). Thus, our study, surveying four rodent species, is not comprehensive enough for addressing rodent relationships in the framework of eutherian systematics. Indeed, the four species sequenced to date represent just three of the ~30 families of rodents. Accordingly, we can consider our interpretations preliminary, pending additional complete mtDNA from other major lineages such as sciurids, castorids, dipodoids, geomyoids, and pedetoids. Only with the analyses of these sequences can more final answers be provided regarding the phylogeny of the order Rodentia.

Acknowledgments

We are indebted to Dr. F. Catzeflis for kindly providing the dormouse tissues (specimen number V-779) and for his valuable comments on the results. We also thank C. Lanave, A. M. D’Erchia, and C. Gissi for their valuable help and M. Longiro for revising the English text. The constructive comments of two anonymous referees have been very much appreciated. This study was supported by grants from MURST (Italy) and the European Commission HCM ERB CHRXCT930254.

LITERATURE CITED


MANOLO GOUY, reviewing editor

Accepted January 16, 1998