Letter to the Editor

Sequence Gaps Join Mice and Men: Phylogenetic Evidence from Deletions in Two Proteins

Celine Poux,* Teun van Rheede,* Ole Madsen,* and Wilfried W. de Jong*†

*Department of Biochemistry, University of Nijmegen, Nijmegen, The Netherlands; and †Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam

Recent nuclear sequence analyses have provided evidence that primates and rodents are more closely related than previously believed (Madsen et al. 2001; Murphy et al. 2001a, 2001b). This proposal is difficult to reconcile with morphological insights (Liu et al. 2001; Novacek 2001) and is not generally supported by current mitochondrial sequence data (Reyes, Pesole, and Saccone 2000; Nikaido et al. 2001; Arnason et al. 2002; Janke et al. 2002). Moreover, the supporting data and analyses have been criticized on methodological grounds (Rosenberg and Kumar 2001). Here we report deletions in two nuclear protein-coding genes that lend independent support to this contested grouping.

Some 18 orders of placental mammals are currently recognized, but their phylogenetic relationships remain highly controversial. Extensive sequence comparisons of mainly nuclear genes support a basal division into four major clades (Xenarthra, Afrottheria, Laurasiatheria, and Euarchontoglires), which has far-reaching implications for early mammalian biogeography and morphological diversification (Murphy et al. 2001b). Euarchontoglires is composed of the orders Primates, Rodentia, Lagomorpha (rabbits, hares, and pikas), Scandentia (tree shrews), and Dermoptera (flying lemurs). In contrast, morphology groups Primates, Scandentia, and Dermoptera with Chiroptera (bats) in the clade Archonta, whereas Rodentia and Lagomorpha (jointly called Glires) are in a distant clade with Macroscelidea (elephant shrews) (Liu et al. 2001; Novacek 2001). Also, sequence data from 12 proteins encoded by the mitochondrial genome generally do not support Euarchontoglires (e.g., Nikaido et al. 2001) or even maintain rodent polyphyley in many cases (Reyes, Pesole, and Saccone 2000; Arnason et al. 2002; Janke et al. 2002). Only by excluding some taxa with high or atypical substitution rates (or both) can sound mitochondrial support be obtained (Waddell, Kishino, and Ota 2001). Establishing the monophyly of the most speciose eutherian order, Rodentia, and finding its sister group has indeed been most difficult to solve on the basis of sequence evidence (e.g., Graur, Hide, and Li 1991; Adkins et al. 2001; Huchon et al. 2002). As for the molecular data sets giving support to Euarchontoglires, it has been questioned whether these are actually able to resolve the relationship of rodents and primates or whether more genes and longer sequences are needed (Rosenberg and Kumar 2001). Given, too, that Euarchontoglires is the least supported of the four major clades in some analyses (Madsen et al. 2001), additional evidence for their monophyly is certainly needed. This could be provided by "rare genomic changes," such as insertions and deletions (indels) in proteins (Rokas and Holland 2000). Indels in protein-coding DNA sequences require more complex mutational mechanisms and are generally more constrained than single base substitutions. Such indels can therefore be good indicators for monophyly, as demonstrated already for two of the other major clades, Xenarthra (van Dijk et al. 1999) and Afrotheria (Madsen et al. 2001), as well as in deeper vertebrate phylogeny (Venkatesh, Erdmann, and Brenner 2001).

While studying genes involved in various neurodegenerative disorders, we noticed two deletions that might be informative for the naturalness of Euarchontoglires. One is a large deletion in exon 8 of the gene for spinocerebellar ataxia 1 (SCAI), resulting in an 18-residue deletion in the encoded protein (fig. 1, top). The other is a 6-bp deletion at the 5' end of the intronless coding region of the prion protein gene (PRNP; fig. 1, bottom). Both deletions perfectly distinguish Euarchontoglires from all other placental and outgroup marsupials. Obviously, the most parsimonious interpretation is that these deletions originated once and independently in the SCAI and PRNP genes of the last common ancestor of Euarchontoglires, thus supporting their monophyly. If the morphological or mitogenomic trees are true, both deletions must have originated at least twice in exactly the same lineages.

Although reversal of the observed deletions in SCAI and PRNP is difficult to imagine, a repeated origin cannot totally be excluded. Indels are certainly not free from homoplasy, especially in regions with sequence repeats. In the SCAI gene, for example, a sequence repeat C TG TCN CCC, coding for Leu-Ser-Pro (underlined in fig. 1, top), might in principle have triggered the large deletion more than once. In the middle of this same region, a 6-bp deletion has caused the loss of two alanines in armadillo, whereas a 3-bp insertion results in an additional alanine in most Laurasiatheria (fig. 1, top). This latter insertion might indeed agree nicely with a basal separation of Eulipotyphla (represented here by hedgehog and mole) from the other Laurasiatheria (Murphy et al. 2001b). However, both the deletion and the insertion are likely to be caused by the GCC (Ala) repeat in this gene region and therefore to have little phylo-
**Fig. 1.**—Deletions in the SCA1 protein (top) and the prion protein gene (bottom) support Euarchontoglires. Protein and DNA sequences, respectively, are shown as being most informative. Sequences correspond with positions 415 to 445 in the human SCA1 protein, and with nucleotides 1–44 of the coding sequence of the human \( \text{PRNP} \) gene. Eutherian species are grouped according to the four recently proposed basal clades of placental mammals (Murphy et al. 2000b). Gray shading emphasizes the overall sequence conservation; -- denotes alignment gaps. The underlined Leu-Ser-Pro repeat in SCA1 is discussed in the text. Most sequences were newly determined by direct sequencing of PCR-amplified genomic DNA fragments and can be found with full species names under accession numbers AJ438463–AJ438487 for SCA1 and AJ438193–AJ438207 for \( \text{PRNP} \). Human and mouse SCA1 sequences are from the database (a, XM004164; b, NM009124), and \( \text{PRNP} \) sequences indicated with c from Wopfner et al. (1999).

**Acknowledgments**

This work was supported by grants from the Netherlands Organisation for Scientific Research and the European Commission.

**LITERATURE CITED**

New information and insights into the evolutionary relationships among mammal species, as presented by various scientific works:


Rodney Honeycutt, reviewing editor

Accepted June 2, 2002