# THE TREE OF LIFE

## Deciphering the puzzle of relationships between mammals using supertrees

THE LONG-STANDING GOAL OF CLASSIFYING life – the science of systematics – is to uncover the Tree of Life, a detailed depiction of the evolutionary history of all life. Although the field of systematics is more than 200 years old, it is only recently that the "molecular revolution," and with it the ever-increasing ability to sequence the DNA of organisms quickly and cheaply, has brought the realization of the Tree of Life tantalizingly close. For instance, the complete nuclear genomes of several mammals are now known (human, chimpanzee, mouse, rat, and dog), with the genome projects of at least 26 other mammal species under way or nearing completion. Even better, the genome of the mitochondrion – small organelles inside the cell that each contain their own, smaller set of DNA (about 16,000 base pairs) separate from that in the nucleus – has been fully sequenced for more than 140 species spanning the breadth of mammals. However, molecular information remains limited outside these "model mammals." No single gene has been sequenced for every mammalian species: the best represented gene, cytochrome *b*, has been completely sequenced for less than one-third of all species. Similarly, as of March 2004, more than 99 percent of the nearly two million Carnivora sequences present in GenBank (an international database for genetic sequence data) were for the dog. With such incomplete information, how is it possible to reconstruct the evolutionary history of all mammals?

The key is to look beyond the DNA to take account of the entire body of information that has been amassed on the evolutionary relationships of a given species. Few species have been fully characterized on a molecular level, but some phylogenetic information exists for all mammals, if only relatively crude taxonomic information, such as the placement of a species in a genus, family, or order. A recent approach called supertree construction involves amalgamating all these fragments of information about a species' relationships into a single all-encompassing tree. Instead of analyzing characters of the organism directly (e.g., DNA sequences or morphological traits), a supertree is pieced together from the trees suggested by the separate lines of evidence. Roughly speaking, supertree construction is akin to building a single picture from multiple, incomplete jigsaw puzzles of that picture. But, because each puzzle overlaps with at least one other one, the supertree can be larger than any of the contributing puzzles, with the overlap showing how they fit together. Puzzles that do not overlap directly can be positioned relative to one another by a series of overlapping puzzles. Also, because the trees being combined need

not agree with each other absolutely, different analytical procedures are used to find the supertree that provides the best fit to the different puzzles.

The combining of trees in a supertree approach currently succeeds where direct analyses of character data fail because different data types often cannot be analyzed using a single analytical procedure. For instance, the evolution of DNA sequences is arguably best examined under a statistical, probabilistic framework that is generally not workable for morphological structures (e.g., dimensions of the skull or the presence or absence of a given feature). We can say how often we expect a specific change in a DNA sequence to occur (on average), but not how often we expect wings to arise within mammals. However, this problem of trying to analyze different types of information simultaneously does not apply when combining the trees that emerge from them. For example, two trees that conclude that mice and rats are more closely related to each other than either are to seals, can always be compared and combined, regardless of the data on which they are based.
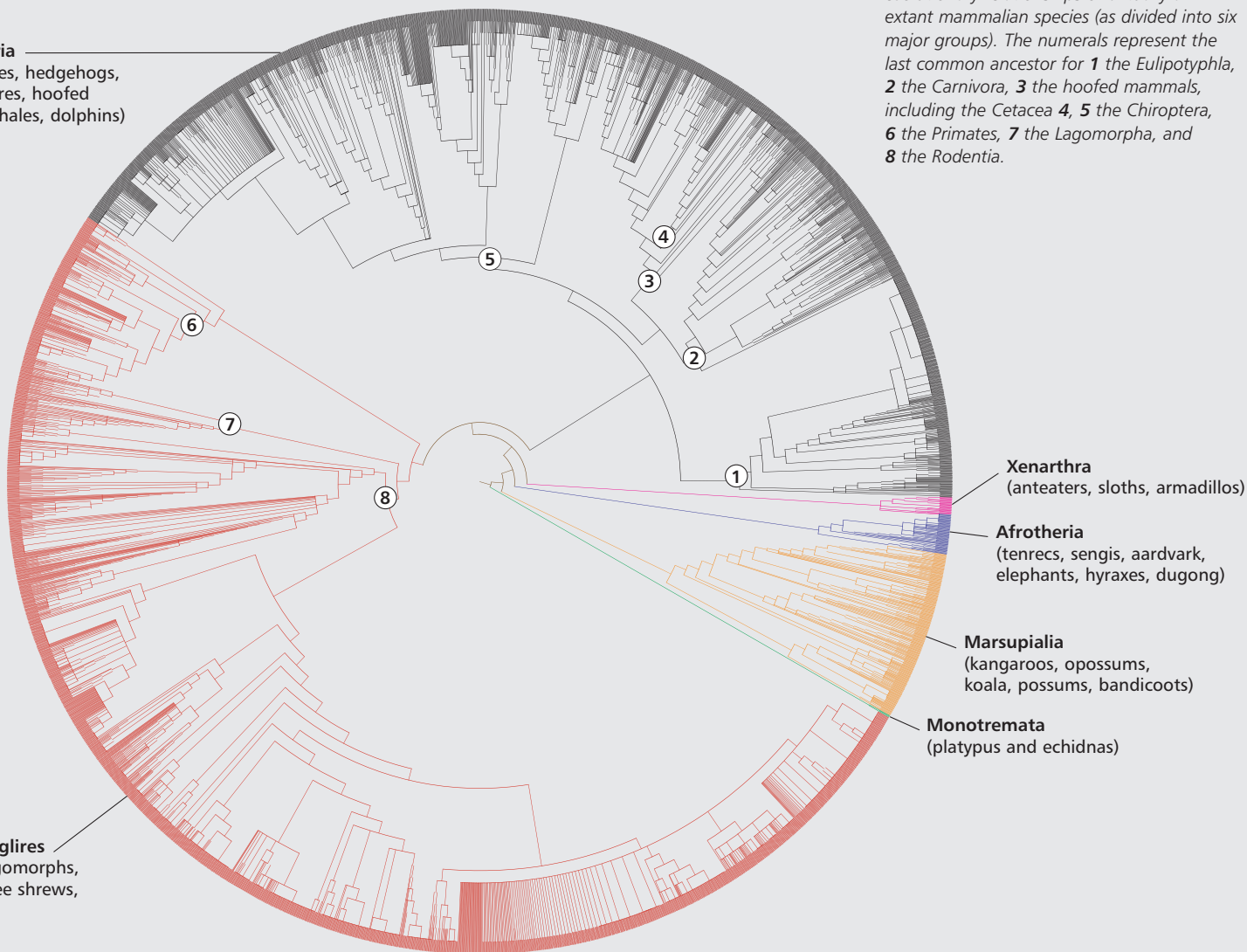
The first attempt at a complete supertree of mammals shows the evolutionary relationships of more than 4,500 extant species, or >98 percent of all extant species (depending on the species list). As such, it represents by far the most complete Tree of Life for mammals yet produced and, because mammals are so well studied compared with other organisms, is one of the largest species-level trees ever assembled. Furthermore, divergence times throughout the tree were estimated by mapping DNA data from 66 genes onto the tree and calibrating them against 30 fossil time-points under a local variant of the molecular-clock hypothesis (which specifies that DNA or proteins evolve at a relatively constant rate). The supertree confirms the growing consensus (based largely on molecular studies) of relationships between the major lineages, but extends this information to the tips of the tree (species) for the first time. The egg-laying monotremes were the first group of mammals to branch from the tree and four major lineages (superorders) of placentals can be identified: the Afro-Arabian Afrotheria; the Euarchontoglires and Laurasiatheria, with roots in the ancient supercontinent of Laurasia; and the South American Xenarthra. The distinct biogeographic distribution of these superorders supports the hypothesis that the drifting of the continents caused by plate tectonics has helped drive evolution within placentals. Within these lineages, the whales (Cetacea) are shown to have evolved within the even-toed ungulates (Artiodactyla), with hippos being their closest living relatives. Insectivora – long the taxonomic

catch-all for any small, brown, furry nonrodent – actually comprises two distinct groups that have seemingly evolved in parallel, the African Afrosoricida (golden moles and tenrecs) and the largely New World, Asian, or northern hemisphere Eulipotyphla (remaining insectivores). However, this represents only one of many apparent instances of morphological convergence between the superorders. The divergence times reveal that most of the main placental lineages (roughly, the orders) originated up to eight million years before the Cretaceous–Tertiary boundary (65 million years ago), but that the diversification of the living members of these groups occurred only after this point (when the dinosaurs and many other forms of life went extinct). The supertree also reveals problematic areas that remain. The evolutionary history of the rodents is poorly resolved, partly because the group is so big (about 2,000 species) that it is hard to study, and has been comparatively little studied among mammals.

The impact of the mammal supertree goes far beyond its purely systematic aspects, where there is great optimism that it can be used to underpin a more complete understanding of mammals. Dating back to Darwin and Huxley, it has been appreciated that species do not represent independent data points for analysis because they are related through evolutionary history. As such, the biology of a species derives from both purely adaptive reasons and because it has been inherited from its ancestors (known as "phylogenetic inertia," although the individual traits usually remain adaptive). Disentangling these two factors naturally requires knowledge of evolutionary relationships and forms the basis of the so-called comparative method in biology. The more complete and well-resolved phylogenies currently offered by supertrees have the potential to increase both the scope of the biological questions that can be asked and the statistical power of the analyses, because comparisons across species are more successful when the data represent all or most of the species of interest. For instance, it is clear that wings have enormous adaptive benefits and affect the overall biology of those organisms possessing them greatly. If we were analyzing the evolution of flight but only had measures of one bat and one bird (and no nonflying vertebrates), then the conclusions would be severely limited. Further, without a phylogeny, it would be hard to know what to compare. Should comparisons be made among a few bats here and few birds there, or which species should be compared within bats? The most convincing results are those based on as many sam-

*Below* A computer analysis reveals the evolutionary relationships of virtually all extant mammalian species (as divided into six major groups). The numerals represent the last common ancestor for **1** the Eulipotyphla, **2** the Carnivora, **3** the hoofed mammals, including the Cetacea **4**, **5** the Chiroptera, **6** the Primates, **7** the Lagomorpha, and **8** the Rodentia.

**Laurasiatheria**
(shrews, moles, hedgehogs, bats, carnivores, hoofed mammals, whales, dolphins)

**Xenarthra**
(anteaters, sloths, armadillos)

**Afrotheria**
(tenrecs, sengis, aardvark, elephants, hyraxes, dugong)

**Marsupialia**
(kangaroos, opossums, koala, possums, bandicoots)

**Monotremata**
(platypus and echidnas)

**Euarchontoglires**
(rodents, lagomorphs, primates, tree shrews, colugos)

ples as possible, with the decision of what to compare with what having been set within a logical, phylogenetic structure.

For these reasons, complete supertrees open the door to tackling problems in biology that have previously been out of reach, and have provided new insights into the pattern and process of evolutionary change. As examples, a picture of the genealogical relatedness of all mammal species has shed light on questions about which species evolved when, about which characteristics are associated with rapid evolutionary change or no change, and about which characteristics are correlated with increases in species numbers. With continuing refinements in supertree analysis, some of the answers are now becoming apparent.

For example, species richness in mammals (that is, the number of species belonging to a cluster on the tree) is significantly correlated with life histories involving rapid reproduction – shorter gestation periods, quicker interbirth intervals, and larger litter sizes. Previously, with incomplete information

it was difficult to be confident whether such patterns were real or partly artifacts arising because parts of the tree were empty (that is, is a group species-poor because it is really species-poor or because it has been incompletely sampled?). Moreover, the same analyses revealed that the biological characteristics influencing species richness differ among groups: whereas population density is more important in Old World monkeys, body size is the key factor in carnivores.

Perhaps the most important task for a complete supertree is to help preserve mammal species and to understand the problems of their conservation. At present, about 25 percent of mammal species are threatened with extinction. Some groups are clearly more threatened than others: Afrosoricida, Artiodactyla, Dermoptera, Eulipotyphla, Hyracoidea, Monotremata, Perissodactyla, and Primates each contain a greater proportion of threatened species than the mammalian average. Understanding the biological and ecological processes that cause some species to be threatened while others

remain comparatively safe is a priority for conservation, and analyses based on supertrees have shed light in this area as well. For instance, it is now known that although geographic range is the most important of the many traits that increase the likelihood of extinction in mammals, group-specific biological characteristics become important additional determinants of extinction risk only for species greater than about 3kg in size.

These exciting examples reveal how having complete phylogenetic information has helped to solve problems that could benefit facets of society ranging from economics to agriculture to conservation to human health (see Beating Diseases in Primates) and that previously seemed lost to evolutionary history when branches of the tree were not represented in our phylogenies. Ultimately, our ability to better understand the complexity of the world's problems and to adapt to rapidly changing environmental issues will depend on our ability to build a more complete and accurate Tree of Life.

OB-E/RB/MC/JLG/RG/KJ/GM/SP/AP