

## Molecular Dating and the Many Different Types of Mammals

Last time, we touched upon the subject of molecular dating in the context of human evolution. In that case, the differences between the animals were small and only four animals were involved, so the analysis was relatively straightforward. Now we will consider a more complicated problem which requires more sophisticated molecular dating techniques: How and when did the shrew like creatures that lived during the age of the dinosaurs become the wide variety of (placental) mammals we see today?

### 1 The Orders of Mammals

Mammals are animals that have fur or hair, are able to regulate their body temperature, and nourish their young with milk. These and other similarities suggest that all mammals share a common ancestry. Other traits are unique to particular groups of mammals, indicating that the animals in these groups have a closer relationship with each other than they do with other mammals. Even though different animals could have acquired similar traits independently (a phenomenon known as convergence) careful consideration of the pattern of variation in many traits has allowed biologists to define divisions within the mammals that probably do define groups with distinctly different heritages.

First of all, all living mammals can be divided into three groups based on their mode of reproduction: **Monotremes** (like the platypus and the echidna) lay eggs; **Marsupials** (like kangaroos and koala bears) give birth to live young which develop inside their mother's pouch, and **Placentals** or **Eutherians** bear live young and use a placenta to support the fetus during gestation. The eutherians are the largest of these groups, and we will focus on these animals in this lecture.

The eutherians can be further subdivided into several groups called **orders**. These orders are briefly described here, starting with the nine largest, which include most of the eutherian mammals (The numbers of species are derived from Corbet and Hill's *A World List of Mammalian Species, Third Edition* (1991))

- **Rodentia**(*Mice, squirrels, guinea pigs and other rodents*)

The roughly 1800 species of rodents have characteristic large, forever-growing incisors, two in the upper jaw and two in the lower.

- **Lagomorpha**(*Rabbits and their relatives*)

The roughly 70 species of lagomorphs have some similarities with rodents, but have enough distinct features that they are treated as a separate group. For example, lagomorphs have four incisors in their upper jaw, not two like rodents.

- **Chiroptera**(*Bats*)

Bats (of which there are about 1000 species) are the only mammals that can actively fly. This adaptation has affected nearly all parts of their anatomy, including their forelimbs which have been transformed into wings.

- **Carnivora**(*Cats, dogs, bears, seals, etc.*)

Most of the carnivorous mammals (except whales) fall in this group (which includes about 240 species). These animals have distinctive “carnissial” teeth. These are not the fangs we usually associate with carnivorous mammals, but a pair of teeth further back in the mouth that act as shears. For reference, these animals are called “carnivores” when referring to their diet, and “carnivorans” when referring to their ancestry.

- **Primates**(*Monkeys, apes, humans, etc*)

The 200 species in this group includes us. Primates have nails instead of claws, and usually also have opposable thumbs.

- **Cetacea**(*Whales and dolphins*)

These mammals are completely aquatic. They have lost their hind-limbs, and their fore-limbs have been converted into flippers. There are about 80 species in this group.

- **Atriodyctyla**(*Pigs, cows, deer, etc.*)

The 200 species of this group are the “even-toed hoofed animals”, since they have either two or four hooves. These animals all have an ankle-bone, the astralagus, which has a distinctive “double-pulley” shape.

- **Perissodactlya**(*Horses, tapirs and rhinos*)

The “odd-toed hoofed animals”, which have one or three hooves. This group only includes about 20 species, quite a bit less than the atriodyctyls.

- **Xenarthra**(*Armadillos, sloths and anteaters*)

A group of rather strange looking creatures, almost exclusively from South America. Xenarthrans are characterized by a general lack of teeth and by distinctive structures in the vertebrae of their lower back.

There are also a number of smaller groups of distinctive mammals, which cannot be assigned to any of the above orders. These animals are placed into their own groups:

- **Proboscidea**(*Elephants*)

Very large land mammals, with tusks and a trunk. The tips of their toes have hoof-like coverings.

- **Hyracoidea**(*Hyraxes*)

Small animals, that look something like a woodchuck, but are clearly different from rodents (They have four incisors in their lower jaw, for one). They have hoof-like nails.

- **Sirenia**(*Manatees and dugongs*)

A group of aquatic mammals distinct from the cetaceans. For example, their teeth and mouth are adapted to graze on aquatic vegetation.

- **Tubulidentata**(*Aardvarks*)

The aardvark eats ants and termites like an anteater, which explains some of their similarities. However, aardvarks do not have xenarthran vertebral processes or other distinctive xenarthran traits, and so are not included in that group.

- **Pholidota**(*Pangolins or scaly anteaters*)

Another ant and termite eater. This animals' most distinctive feature are the large "scales" covering much of its body.

- **Dermoptera**(*Flying lemurs*)

These creatures are not lemurs nor can they fly. They can only glide on a furry membrane extending between their limbs and their body.

Finally, we have one more group of animals which historically belong to an order called **Insectivora**. These are small mammals that don't fit in any of the other groups, such as hedgehogs, moles and shrews of various sorts. Since there are no unique features which are common to all the mammals in this group, it has long been recognized that this "order" may include more than one distinct group of animals. In particular, two groups of shrews, the elephant shrews (which have a distinctive snout) and the somewhat squirrel-like tree shrews have been given their own orders, called **Macroscelidea** and **Scandentia**, respectively.

## 2 Arranging the orders with morphology

While the eutherian mammals show sufficient commonalities to justify the idea that they have a common ancestry, and the orders do appear to be plausible subdivisions (with the exception of insectivora), the relationships between the different orders is far from clear. There are no traits in living mammals that clearly link any two orders together into a larger group within the eutherians. Even using the fossil record is not as helpful as one might hope. There are fossil representatives of all the orders (see figure 1), but few fossils clearly document a close relationship between two orders. A noteworthy exception, however, is provided by a recently discovered ancestor of the whales which has a "double pulley" astralagus in its heel, indicating a close relationship between cetaceans and artiodactyls.

Relationships between the orders have been inferred by comparing rather subtle morphological (physical) characteristics. The results of one such analysis are shown in figure 1. The close relationship between lagomorphs and rodents, or the one between bats and primates, appears rather often in these analyses, and these groups have even been given the names **Glires** and **Archonta**. However, very few of these relationships are strongly supported.

The available fossil evidence does also not place very tight constraints on when these various orders separated from one another. The fossil record shows that members of several orders are found soon after 65 million years ago, when the last of the

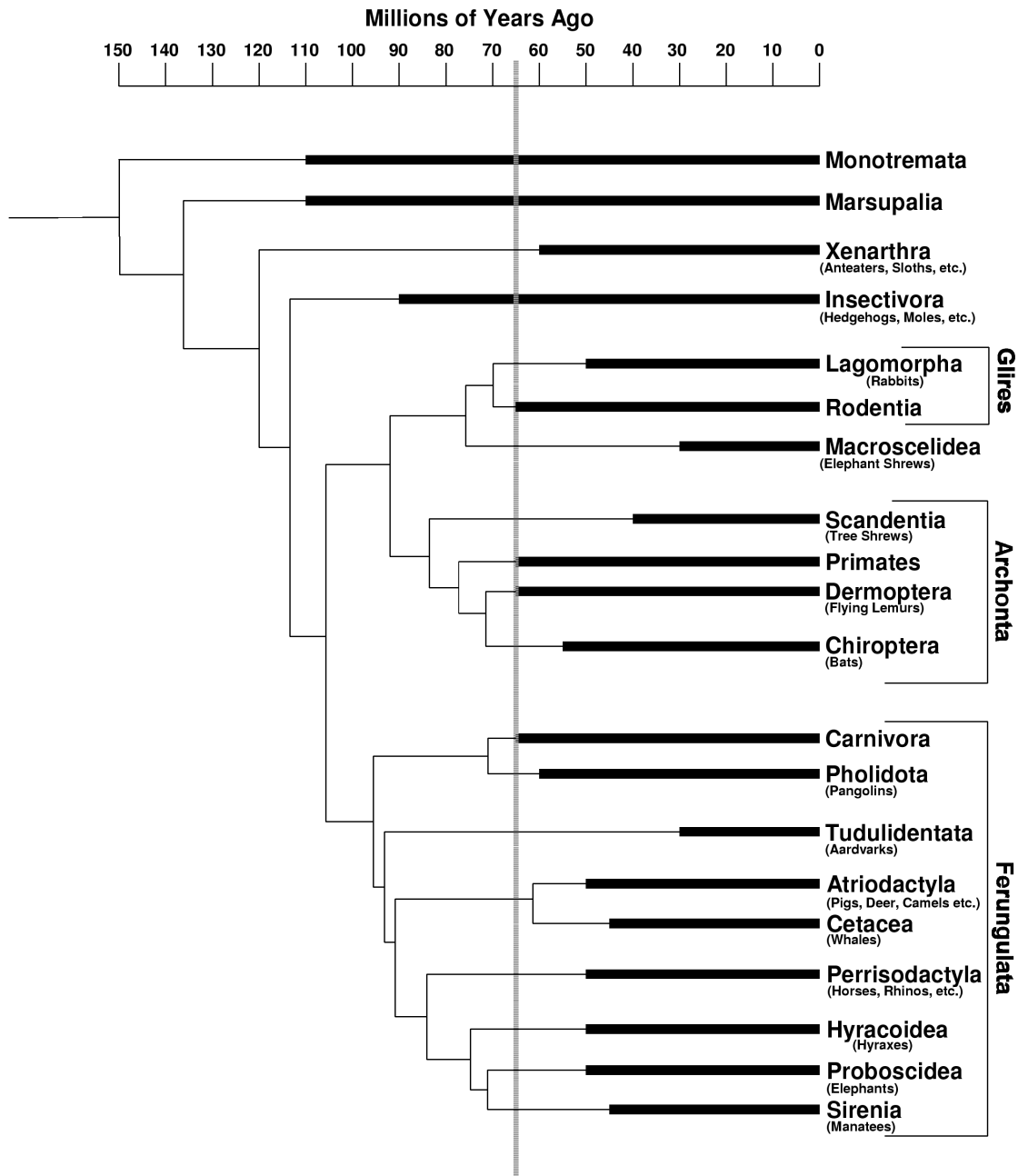


Figure 1: The fossil record and relationships of the eutherian mammal orders based on morphological (physical) characteristics. From Shoshani and McKenna "Higher taxonomic relationships..." in *Molecular Phylogenetics and Evolution* Volume 9 (1998), pp 572-584. The thick lines indicate how far back the fossil record of each group extends (breaks in these records and groups only known from fossils are not shown). The thin lines indicate the relationships between these orders based on comparing the morphological traits in these organisms. *Note that the timing of the various forks in the tree is rather arbitrary, since without fossil evidence, there is no way to determine when the various groups branched from one another based on only morphological characters.* Indeed, based on the fact that most groups only date back to 65 million years ago (indicated by the vertical gray line), when the dinosaurs died out, it has been suggested that most of these groups diverged around that time.

giant dinosaurs died out. This suggests that the different orders may have diverged when the eutherian mammals diversified to fill the void left by the dinosaurs. However, without additional information, it is difficult to confirm or deny this idea. Data from DNA sequences can potentially shed important light on this situation.

### 3 Molecular Methods

As we saw in the last lecture, as DNA is passed from one generation of animals to the next, mutations (changes to the DNA sequence) accumulate over time. These mutations are easier to quantify than changes in the physical characteristics of the animals, so it should be more straightforward to determine relationships between the various groups. Furthermore, mutations in the DNA sequence could accumulate in a more regular way than changes in physical characteristics, so these data might provide a means to estimate when the various groups diverged from one another.

The raw data for this analysis are a collection of comparable DNA sequences obtained from a representative variety of living eutherian mammals (at least one from each order). Each DNA sequence is compared with all of the others to find all places where one of the nucleotides has been replaced by another (that is, a point substitution mutation has occurred, as when ...ATGC... becomes ...ATTC... ). Last time we found it was possible to deduce relationships between the great apes by simply counting the number of differences between any two sequences and comparing the resulting numbers. Untangling the relationships between different mammalian orders, on the other hand, requires more sophisticated techniques.

#### 3.1 Variations in mutation accumulation rates

When using non-coding sequences of closely related species, it is reasonable to suppose that mutations could accumulate at a constant rate. However, to establish the relationships between the orders of eutherian mammals we must consider a much larger variety of animals, which have had a much longer time to accumulate distinct mutations. Furthermore, the only DNA sequences that have been obtained for all the different orders are from genes (pieces of DNA that contain information for making proteins), which are easier to identify than non-coding sequences but are less likely to accumulate mutations in a regular, standard way. We therefore cannot assume that the DNA sequences will accumulate mutations at the same rate in all animals at all times.

Indeed, there is direct evidence that the mutation accumulation rate is different for different animals (even in non-coding regions). The mutation accumulation rate can be estimated in certain cases by comparing DNA sequences from two animals whose lineages diverged a known amount of time ago (based on the fossil record). Figure 2 shows some estimates of the mutation accumulation rate of a particular stretch of DNA of different animals as a function of their body size. Larger animals seem to accumulate mutations more slowly than small animals, and cold-blooded creatures accumulate mutations more slowly than warm blooded creatures. These and other trends observed in the mutation accumulation rates of different animals are still not fully understood. Fortunately, even if mutation rates are accumulating at different

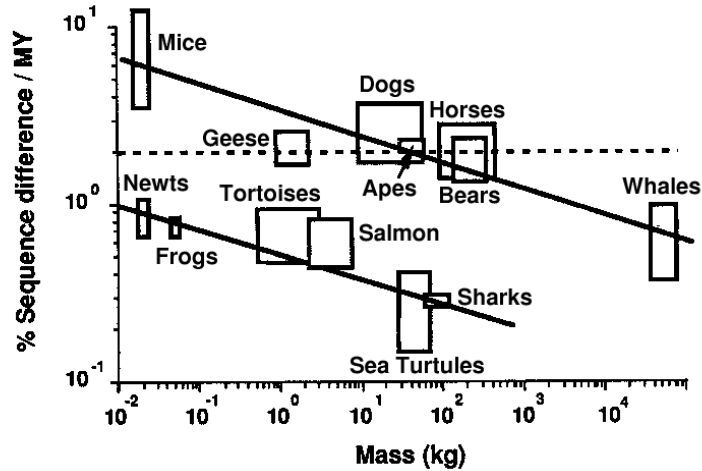
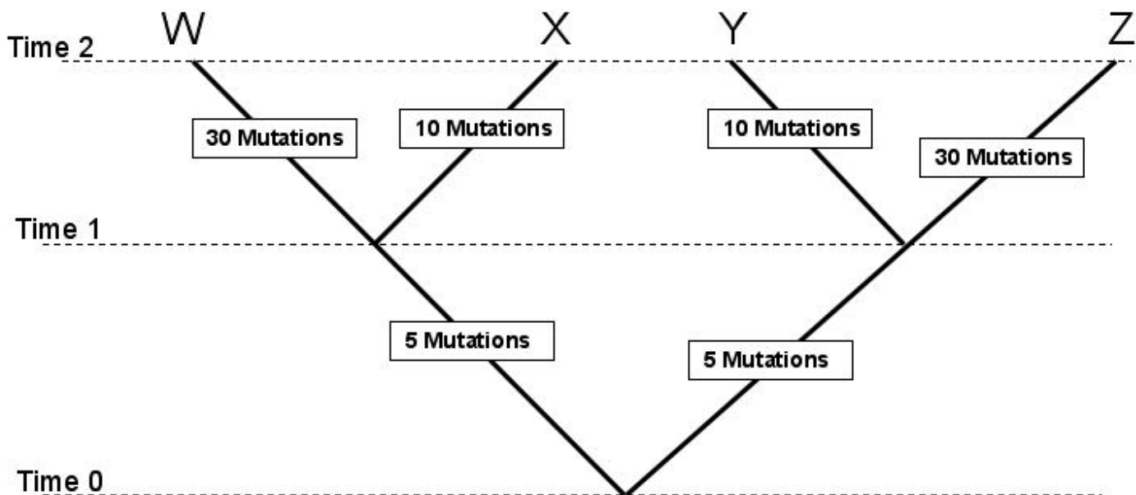


Figure 2: Different mutation accumulation rates in different animals. This figure is from an article by A.P. Martin and S.R. Palumbi in the *Proceedings of the National Academy of Sciences USA*, Volume 90 (1993) pg 4087. The graph shows the mutation accumulation rate in mitochondrial DNA of different animals as a function of the body size of the animal. Each box corresponds to a particular animal, the dashed line indicates where the boxes should lie in all animals accumulated mutations at the same rate. Note that larger animals have lower rates, and cold-blooded animals have lower rates than warm-blooded animals.

rates in different animals, we can still derive relationships and even estimate the times when various groups diverged from one another.

### 3.2 Making family trees with varying rates

In order to demonstrate that it is possible to establish relationships between the orders and construct a family tree of eutherian mammals without assuming that mutations accumulate at a constant rate, consider the following simple situation. We have four animals, named W, X, Y and Z, living today (also called time 2). these animals are all descended from a single animal that lived at time 0, as shown here.



At time 0, two lineages of descendants from this ancestor begin to diverge, accumulating 5 mutations each by time 1. At time 1, both these lines split again, forming the four lineages that lead to the four animals. Between time 1 and time 2, the lineages leading to W and Z accumulate 30 mutations, while those leading to X and Y accumulate 10 mutations.

Let us say we analyze the DNA of the four animals and attempt to figure out how these animals are related. Comparing the DNA of animals X and Y, we find they have 30 mutational differences between them (15 from X, 15 from Y). Calculating the number of differences between each pair, we get the following table

	X	Y	Z
W	40	50	70
X		30	50
Y			40

Looking at this table, we see X and Y have the smallest number of differences between them. If we assumed these animals accumulated mutations at a constant rate, then we would conclude that X and Y are the most closely related. Of course this is incorrect, X is more closely related to W than Y. Fortunately, the other numbers in the table allow us to figure out that different lineages accumulated mutations at different rates, and determine the true relationships between these animals.

Look at the first row of the table. It says that W has 40 differences from X and 50 differences from Y. If X and Y were really the animals with the most recent common ancestor, then this would imply Y accumulated mutations more rapidly than X. Now look at the last column of the table. This says Z has 50 differences from X and 40 differences from Y. This would mean X accumulated mutations more rapidly than Y if X and Y have the most recent common ancestor. These data therefore show that there is a contradiction if we assume X and Y have the most recent common ancestor.

Such a contradiction does not arise if we assume (correctly) that W and X are the ones with the most recent common ancestor. Y has 30 differences from X and 50 from W, implying W accumulated mutations faster than X. Z has 50 differences from X and 70 differences from W, which also indicates W accumulated mutations more quickly. Thus, by considering all of the differences, it is possible to uncover proper relationships even when the mutation accumulation rate varies between different lineages.

Applying such considerations to the analysis of real DNA data requires computer programs to find the most likely tree efficiently. In practice, several different methods exist to determine relationships when the mutation accumulation rate is not constant. Some use the total number of differences between groups (as in the example above), and are called **distance methods**. Other methods look at individual mutations. For example, between time 0 and time 1, the common ancestors of W and X accumulated 5 mutations, and so W and X share five identical mutations. This would be relatively unlikely if the ancestors of W and X accumulated mutations independently. Therefore we could determine that W and X are closely related because it reduces the number of “coincidental” mutations. This is the basis of **maximum parsimony** and **maximum likelihood** methods. All three of these techniques have advantages and disadvantages, which we do not have time to discuss here.

A recent example of the relationships between the orders derived from molecular data is shown in figure 3 (other analyses yield similar results). This analysis shows

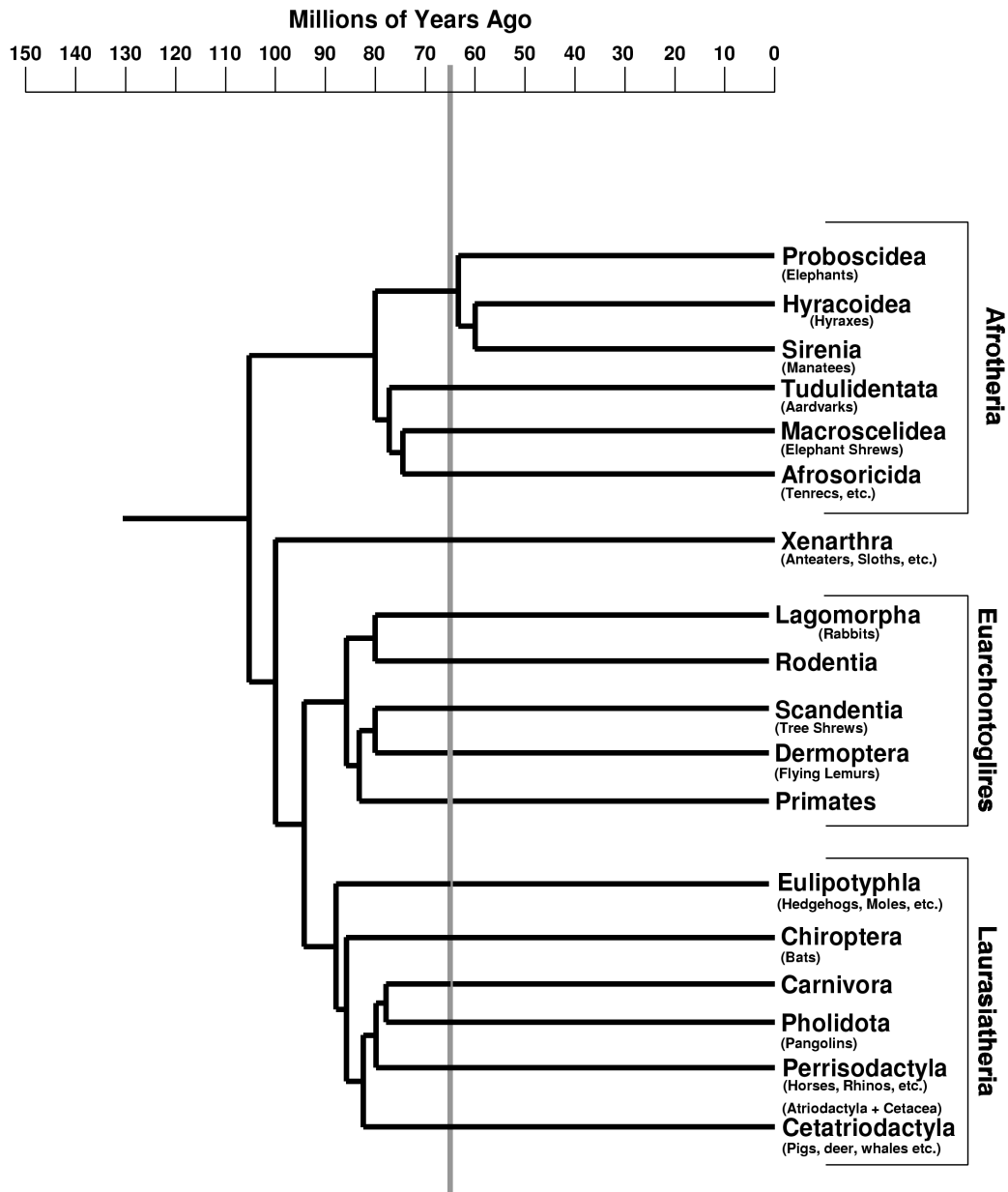


Figure 3: The relationships of Eutherian mammals based on molecular data. From Springer et al's article in *Proceedings of the National Academy of Sciences* Vol 100 (2003) pp 1056-1061. The relationships derived in this way are not exactly the same as those shown in figure 1. The estimates of when these various groups diverged based on the Bayesian statistical techniques discussed in the text, and indicate the orders diverged well before 65 million years, when the last of the dinosaurs died out. (Note that Cetacea and Atriiodactyla are fused into a new order, this is because whales are more closely related to certain atriiodactyls like hippos than hippos are to other atriiodactyls like pigs.)



the different orders can be lumped together into four larger groups:

- **Afrotheria** which includes elephants, aardvarks and manatees. As its name implies, most of these animals are found in Africa.
- **Xenarthra** This order of South American animals is in a class by itself
- **Euarchontoglires** This group includes rodents, primates and a number of smaller groups.
- **Laurasiatheria** The most diverse group, including both the even and the odd toed hoofed animals, the carnivorans, the whales and the bats.

This arrangement of the mammalian orders does not perfectly match the pattern based on morphological characters (compare figures 1 and 3. However, these relationships are not implausible, and do show some interesting geographical correlations. Afrotherians are found mostly in Africa, while Xenarthrans are found mostly in South America. Euarchontoglires and Laurasiatherians are found all over the world today, but their earliest representatives of these groups appear to be from the northern continents. These correlations will become even more interesting when we consider the timing of the various branch points in this tree.

### 3.3 Finding time with Bayesian statistics

Once the particular branching pattern of the family tree (what is known as the tree topology) has been established, we can attempt to estimate when various groups branched off from one another. If we could assume that the mutation accumulate rate was the same for all parts of the tree, this would simply require using the fossil record to estimate the mutation accumulation rate. Then we could convert the number of mutations between any two animals into a time when their lineages diverged from one another. In reality, different lineages probably accumulated mutations at different rates, so this simple method will not work and it may even appear that it will be impossible to estimate reasonable divergence times. Fortunately the situation is not quite so bleak because the sequence data does provides information about relative mutation accumulation rates in different lineages. Recall that in the example above we were able to establish that animal W's lineage accumulated mutations faster than animal X's lineage based only on intercomparisons between the four animals. Therefore a careful analysis of the DNA sequences should impose constraints on the mutation accumulation rates in the various branches, as well as the age of the various branching points in the family tree.

If the mutation rate for each branch was known, then it would be relatively easy to calculate what sort of differences between the various animals should be observed today. On the other hand, going the other way and inferring the rates from the observed differences is not so straightforward. Any given set of differences *could* have been produced by many different possible variations in the mutation accumulation rate, and we must figure out which of these possibilities is most likely what *actually* occurred.

To illustrate how such problems involving the relative likelihoods of different possibilities can be solved, let us consider a simpler, more down-to-earth problem: A

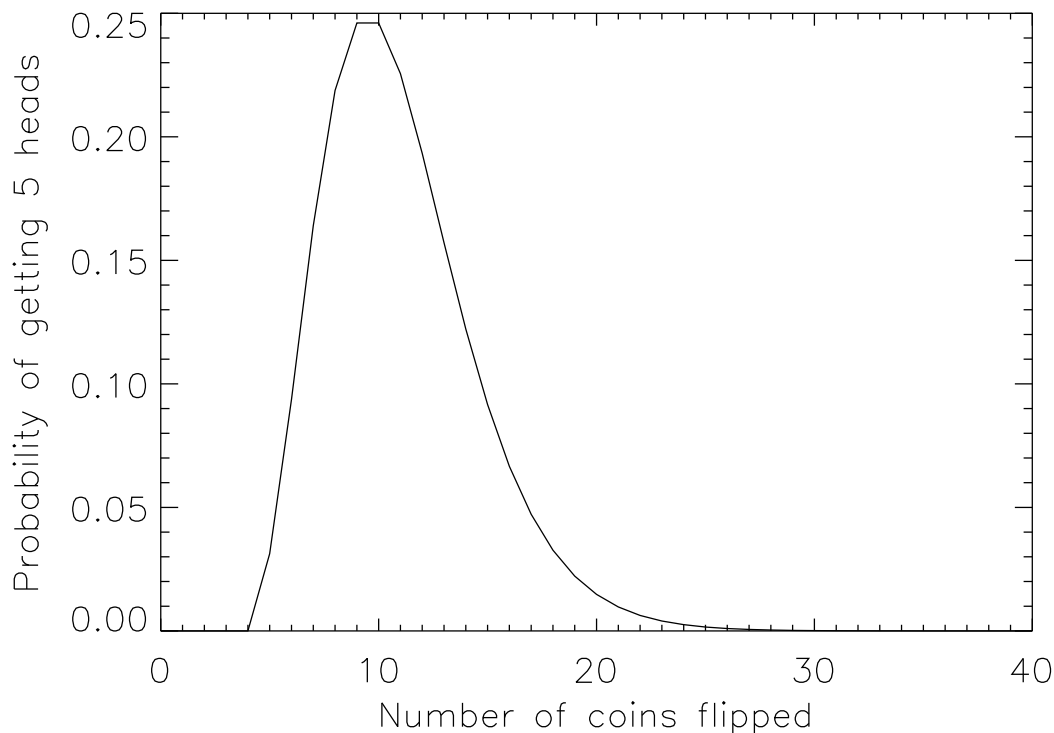


Figure 4: The probability of obtaining five heads by flipping a coin as a function of the number of times the coin is flipped. If the coin could have been flipped any number of times with the same probability, then the curve giving the probability that the coin was actually flipped that number of times given it came up heads five times has the same shape.

coin is flipped some unknown number of times, and we are told **only** that the coin came up heads five times. How many times was the coin flipped? Five heads could be obtained with 5 flips, or 10 flips, or 100 flips, so again we have multiple ways this result could have happened. The problem is to calculate the likelihood that any of these results was actually what happened.

For every possible number of coin flips, we can easily calculate the probability that the coin comes up heads 5 times (see figure 4). There is no chance of getting 5 heads until the coin has been flipped five times. With five flips, we would need it to come up heads every time, which is rather unlikely. The probability of getting the desired number of heads increases with number of flips until we hit a peak around 10 coin flips. This makes sense, since you expect to get about half the coin flips as heads. The probability of getting the five heads decreases again when the number of flips is very large, since it is very unlikely that only five flips will come up heads.

These considerations give us the probability of getting the desired number of heads given the number of flips. However to estimate the probability that the coin was flipped a certain number of times requires additional assumptions. Say we assume that the coin could have been flipped any number of times with equal probability, then we can calculate the probability that the coin was flipped  $x$  times and that it came up

heads five times. This is just the product of the probability that the coin was flipped  $x$  times (which is the same for every value of  $x$ ), and the probability of getting five heads given the coin was flipped  $x$  times (illustrated in figure 4). The resulting curve has the exact same shape as the curve in figure 4, which has a maximum around 10 coin flips. Therefore, assuming the coin *could have been* flipped any number of times, we can conclude based on the number of heads that it most likely was *actually* flipped about 10 times.

This is an example of **Bayesian statistics** or **Likelihood analysis**, and the same sort of idea can be applied to the molecular data. We can calculate the probability of getting the observed differences given every possible combination of mutation accumulation rates. Then, assuming the actual rates could have been any value within a wide range, we can compute the most likely values for the actual rates. (In practice, the rates are limited to “reasonable” values). The constraints on the timing of certain branch points imposed by fossil data can also be included in these analyses.

These calculations can yield estimates of the numbers of mutations accumulated in each branch and when the various branch points occurred. One example of the results of such an analysis is shown in figure 3. Interestingly, it shows that the various groups of mammals diverged for each other well before 65 million years ago, well before the giant dinosaurs died out. This does not necessarily mean that we had mammals that looked like horses or monkeys roaming around with the dinosaurs, and indeed there is no fossil evidence that such creatures existed. However, it may imply the ancestors of these animals were separated from one other well before they acquired the traits that distinguish them today.

Recall that the four major groups seems to be correlated with geography. Afrotherians are from Africa, Xenarthrans are from South America, and the other two groups are from northern continents. The Afrotherians split from the others about 100 million years ago. This was just about when Africa and South America first began to split apart due to continental drift. Therefore we can imagine that there was a particular group of primitive eutherians spread between Africa and South America. When these continents split apart, the two groups of mammals went their separate ways. Sometime later, some animals from South America found their way into the northern continents of North America and Eurasia. Once the dinosaurs died out 65 million years ago, all these mammals diversified on their separate continents and eventually adopted the distinctive characteristics we see today.

With this lecture, we must end our discussions of when things happened in the history of life on earth and turn to events in outer space. In particular, next time we will discuss meteorites and the information they contain on how and when the solar system formed.

## 4 References

For a nice popular summary of today's mammals, see

- E Gould and G McKay *Encyclopedia of Mammals* (Academic Press 1998)

For the fossil evidence for the evolution of mammals, see the relevant chapters in:

- E.H. Colbert, M. Morales and E.C. Minkoff *Colbert's Evolution of the Vertebrates* (J Wiley and Sons 2001)
- Robert Carroll *Vertebrate Paleontology and Evolution* (W.H. Freeman and Company 1988)

For the basics of genetics, a good place to start is

- Larry Gonick and Mark Wheeler *The Cartoon Guide to Genetics* (Perennial Press, 1991)

Some books on reconstructing relationships from genetic data at the college level, see

- Wen-Hsiung Li *Molecular Evolution* (Sinauer 1997)
- M Nei and S Kumar *Molecular Evolution and Phylogenetics* (Oxford U Press 2000)

For the latest work on dating the divergences of eutherian mammals with molecular techniques, see:

- M Springer et al "Placental mammal diversification and the Cretaceous-Tertiary boundary" in *Proceedings of the National Academy of Sciences (USA)* Vol 100 (2003) pp 1056-1061.
- M Hasegawa et al "Times scale of eutherian evolution ...." in *Genes and Genetic Systematics* vol 78 (2003), 267-283.