Chapter 4: Relatedness and taxonomy

Even since Darwin, taxonomy, the framework of formal names used for scientific communication, has acquired a clear and definable mission of representing evolutionary relationships. However, it took another hundred years for the implications of that mission to become clear. First it requires that one understand evolutionary relatedness in terms of common ancestry: the more recently two organisms share a common ancestor, the more closely related they are. Then, because common ancestry is captured in phylogenetic trees, only classification systems that mirror the tree of life faithfully represent evolutionary relatedness. Classification, relatedness, and trees are really three sides of the same coin, so to speak. In this chapter we begin by clarifying a formal understanding of relatedness that applies equally to family trees (pedigrees) and phylogenetic trees. Then we discuss the common confusion of relatedness versus similarity and explain why modern classifications focus on phylogenetic relatedness rather than similarity.

The concept of relatedness

The concept of relatedness applied to the branches of a phylogenetic tree mirrors that used in discussions of human familial relationships. In both cases, the degree of relatedness of two living organisms is dictated by how many generations earlier they last shared a common ancestor. Thus, the following two statements should be interpreted in parallel: (1) you are more closely related to your siblings than to your first cousins, and (2) you are more closely related to chimpanzees than to mice.

Starting with you and your siblings vs. first cousins, what does the statement mean? When we pose this question to classes, the commonest first response is that you share more genes in common with your siblings than with your first cousin. While this is probably true, is it the true basis of the closer relationship or just a consequence of the closer relationship? Suppose that through a freak of genetic segregation you actually shared more genes in common with your first cousin than with a sister (we will leave you to work through how this improbable event might happen), would you now say that you are more closely related to your first cousin than to your sister? No, of course not. Relatedness is about kinship and ancestry, not about the actual assortment of genetic material.

The real measure of relatedness is not genetic similarity, nor any other kind of similarity, but recency of common ancestry. You are more closely related to a sibling than to a first cousin because you share more recent common ancestors with your sibling (your parents) than you do with your first cousin (your grandparents). Likewise, you share more recent common ancestors with your first cousins (grandparents), than you do with your second cousins (great grandparents).

If degrees of relatedness in a pedigree are unfamiliar, you might want to draw out a portion of your family tree and work on finding common ancestors until you are comfortable with the concepts. As an example, the figure shows the male descendants of famous biologist and Darwin contemporary, Thomas Henry Huxley. By focusing only on
the male line, the pedigree is tree-like and easier to follow. You could equally follow the female line to obtain a tree like pedigree, but the historical documentation in human populations tends to make this harder. To emphasize the parallels to phylogeny, where we are usually dealing only with living species, consider only the individuals in the lowermost (most recent) generation shown.

This diagram includes Anthony, his brother (Francis), their two male first cousins (Matthew and Stewart), and one of their (perhaps many) male second cousins (Charles). Anthony is more closely related to Francis than to anybody else in his generation, because they share the most recent common ancestor (their father Julian Huxley, who happens also to have been an evolutionary biologist). Likewise, Anthony is more closely related to his first cousins, Matthew and Stewart, than to his second cousin, Charles, because the last common ancestor he shares with his first cousins is their grandfather (Leonard), but the last common ancestor he shares with his second cousin is their great-grandfather (Thomas Henry).

You may have noticed that the names of the parents and grandparents are not needed to determine degree of relatedness – all that matters is the ancestry of the lowermost (most recent) generation. The following figure, therefore, provides all the relevant information needed to figure out the degree of relationships between members of this generation. Given this tree some three-way statements of relationships are true, for example, “Stewart is more closely related to Anthony than to Charles,” and others are false, for example, “Francis is more closely related to Matthew than to Anthony.” If these are challenging, you can practice by listing and scoring as true or false, some of the 60 possible three-way statements involving the five individuals in the most recent generation.
Relatedness and phylogenetic trees

Now we will apply exactly the same principle to species relationships. Why is it true to say that a human is more closely related to a chimpanzee than to a mouse? Because humans share a more recent common ancestor with the chimpanzee (ca. 6 Ma [Mega annae = million years ago], than with the mouse (ca. 75 Ma). Similarly, humans share a more recent common ancestor with a mouse than with a frog (ca. 350 Ma). Thus humans are more closely related to a mouse than to a frog.

<table>
<thead>
<tr>
<th>Species</th>
<th>Last common ancestry with human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frog</td>
<td>350 Ma</td>
</tr>
<tr>
<td>Mouse</td>
<td>75 Ma</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>6 Ma</td>
</tr>
</tbody>
</table>

Given that the times that common ancestors lived can be difficult to determine, and also difficult to memorize en masse, it is fortunate that all the information needed to evaluate degrees of relatedness is present in a tree diagram. A simplified figure can thus summarize the relationships without needing to include actual dates. By convention (Chapter 2), the tree is usually drawn with the recent at the top and the past at the bottom.

Reading relationships from trees may seem challenging at first, so let’s practice using the tree below, remembering again that relatedness is defined in terms of recency of common ancestry. Looking at the tree below, is a salamander more closely related to a human or a lungfish?
Commonly, as many as three quarters of respondents will first answer “lungfish” when posed a question such as this. But, actually, a salamander is more closely related to a human than to a lungfish, because it shares a more recent common ancestor with a human than with a lungfish.

The best strategy for correctly answering the question is to work down the tree to find the point that corresponds to the most recent common ancestor of a salamander and a lungfish. This ancestor is situated at the node labeled $a$. Now do the same thing to find the last common ancestor of a salamander and a human, which is at the node labeled $b$. Because, $b$ is a descendant of $a$ (to get from the root to $b$ you need to pass through $a$), $b$ must have lived after $a$. Therefore, a salamander is more closely related to a human than to lungfish. It might help to put arrows on all branches that point away from the root, and thus from parents to offspring. This may make it clearer that $b$ is a descendant of $a$.

Given the large number of individuals who find such problems difficult, it will be instructive to explore some of the likely causes of confusion. Even if you got the answer right, you will probably find it useful to learn more about how trees are liable to be misunderstood (cite the new paper on this).
The commonest error is to look only at the tips, and focus on the proximity of labels to one another. If you do this you would “see” the tree as the order list: lungfish-salamander-crocodile-mouse-human. Because the salamander is directly next to the lungfish but relatively far away from the human, you might think it is more closely related to the lungfish. Looking along the tips like this will often lead you astray. It is important, therefore, to train your brain to not even see the order of tips, but instead focus on the branching pattern.

To see why the ordering of tips is not a good guide to relationships recall the way that a phylogeny grows by ancestral lineages splitting. It is arbitrary which descendant lineage one draws to the right or left. Thus, two trees with different tip orders can tell the same history (see Chapter 2). For example the tree below shows one of the many ways of rearranging the tips on this tree without changing either the tree or the implied relationships. While the order of tips has changed, neither the tree nor the implied relationships have changed: \( b \) (the last common ancestor of salamanders and humans) is still a descendant of \( a \) (the last common ancestor of salamanders and lungfish).

Conversely, two trees with the same ordering of tips can imply quite different relationships. The tree below retains the tip order from the original tree but shows some truly bizarre evolutionary relationships. For example, a mouse is now more closely related to a crocodile than to a human. If this messed-up tree were correct, a salamander would indeed be more closely related to a lungfish (ancestor \( a \)) than to a human (ancestor \( b \)) - but it is not!
A second source of confusion related to the first, is the tendency to view evolution as a progressive story with humans at the summit. This ladder of life (or Scala Naturae) metaphor, has a long history in human culture, only being rejected with the acceptance of the Darwinian evolutionary model (see Chapter 1). Within the ladder framework, you might reason that lungfish gave rise to salamanders, which gave rise to reptiles (like crocodiles), which gave rise to “lower” mammals (like mice), which then gave rise to the “highest” mammal, humans. Therefore, you might think that a salamander is just one rung up from a lungfish, but three rungs down from a human, and is, therefore, closer to the former.

The problem with ladder thinking is that it thoroughly misrepresents evolution. Salamanders are alive today and none of them are ancestors of humans (or any of the other tips). While it is likely that the last common ancestor of a salamander and a human looked (to a human eye) rather more like a salamander than a human, it was neither a living human nor a living salamander. The last common ancestor of you and your first cousin is your grandparent. Even in the improbable case that your cousin looked exactly like your grandmother, you would not be descended from your cousin! Avoiding ladder-thinking in all its subtle forms (e.g., the use of terms like “lower animals” or “higher
plants”) is one of the greatest and most important challenges in developing a clear understanding of evolutionary biology.

A third kind of error that can lead you astray is “node-counting.” This is where you notice that to go from the salamander tip to the lungfish tip you need only pass through two internal nodes: \( a \) and \( b \). In contrast, to get from the salamander to the human you need to pass through three nodes: \( b \) and the two nodes above. By this reasoning you might conclude that a salamander is separated from a lungfish by fewer speciation events than separate it from a human, and therefore that it is closer to a lungfish.

Recalling that the form of a tree does not change when tips are pruned off (Chapter 2), the number of taxa that are included and happen to have nodes on the path between salamander and human is variable and uninformative. Consider the two trees below, both have the same implications for the relationships of salamanders to humans and lungfish (notice that node \( b \) is a descendent of \( a \) in both cases), but node counting will yield quite different answers.

The last common cause of confusion is a natural tendency to conflate relatedness and similarity. You might believe that the similarities of a fish and a salamander (slimy skin, lack of hair, aquatic reproduction, etc.) are more numerous than the similarities of salamanders and humans (four limbs, lungs). And if you did, you might conclude that salamanders and lungfish are therefore more closely related. To make this mistake would be to conflate ancestry with similarity, two entirely different concepts (see chapter 3).

While closely related organisms usually look quite similar, sometimes they may not be. This is well illustrated by the following example.
This tree shows that crocodiles are more closely related to birds than to lizards. But to the human eye a crocodile seems to have many more features in common with a lizard than to a bird. The branch lengths in the phylogram to the right (see Chapter 3) are hypothetical but resemble those one might expect if branches were drawn proportional to the number of evolutionary changes in physical attributes. This presumably reflects major evolutionary changes associated with the evolution of powered flight. As a result, the list of the physical traits shared by crocodiles and lizards, but absent in birds is longer than the list of traits shared by crocodiles and birds but absent in lizards. In the former category are such features as quadrupedal locomotion, elongated tails, teeth, flat sternum, scaly body, “cold” blood, solid bones, separate collar bones, and many more. Whereas, the latter list would be quiet short: for example it would include the behavior of building and defending nests and a skull with an antorbital fenestra (a hole in front of the eye). Does the larger number of crocodile-lizard similarities make these more closely related?

There are two levels at which we need to answer this question. First, do these traits call into doubt the truth of the tree diagram? The answer is “no.” The features that seem to unite lizards and crocodiles all trace back to the last common ancestor of lizards, crocodiles, and birds (mammals too, actually). They are shared ancestral traits (or plesiomorphies). These traits can be explained on this tree parsimoniously by a single change in state: the lizard and crocodile lineages have retained the ancestral traits, with a single transition for each trait occurring on the lineage leading to the birds. This is illustrated for the tail trait to the right. The shared ancestral traits of crocodiles and lizards do not contradict this tree, though also they do not actively support the tree.

The traits shared by birds and crocodiles are shared derived states, or synapomorphies. These too can be explained on this tree with a single change in state and are, thus, fully compatible with tree. This is illustrated with the antorbital fenestra (AOF). However, unlike the shared primitive states, the shared derived states actively support this tree because, if crocodiles and lizards were sisters, these characters could not be explained with only a single change in state.
Now that we have established that the numerous shared traits of crocodiles and lizards do not challenge the veracity of the tree, we can ask the second question. Accepting the tree, do the many similarities of lizards and crocodiles make them more closely related to each other than to birds? The answer is a resounding “no.” The tree depicts common ancestry, the very basis of “degree of relatedness.” It would not matter how many derived traits evolved on the bird lineage, the crocodile lineage would still share a more recent common ancestor with birds than with lizards. Relatedness is defined based on descent not similarity.

**Taxonomy and Phylogeny**

In the late 18\textsuperscript{th} and early 19\textsuperscript{th} century, naturalists discovered that if one classified organisms based on their traits, a tidily nested taxonomy usually emerged. For example, the vertebrates could be broken up into those with four legs and lungs (tetrapods), most of which lived on land, and those without limbs, which lived in water (fish). Tetrapods, in turn, could be divided into those with an amphibious life style, and those with an amniotic egg that can develop out of water (amniotes). Amniotes, in turn, could be divided into those with feathers and wings (birds), those with fur and milk (mammals), and those with neither feathers nor fur (reptiles). The system was cleanly nested: there were no feathered animals that lacked an amniotic egg, no furred animals that lacked limbs, etc. This can be summarized in a Venn diagram.

![Venn diagram](image)

Why is the nested structure of taxonomy significant? For contrast, consider the best efforts to arrive at a hierarchical classification of books in a library. The Dewey Decimal system first divides books into 10 broad categories including, Language (400 series), Natural Sciences & Mathematics (500 series), and Geography & History (900 series). Below we list some entries under each of these three broad categories. This can also be drawn as a Venn diagram.

<table>
<thead>
<tr>
<th>420-489 European languages</th>
<th>558 Earth sciences of South America</th>
</tr>
</thead>
<tbody>
<tr>
<td>491-495 Asian languages</td>
<td>914 Geography of Europe</td>
</tr>
<tr>
<td>496 African languages</td>
<td>915 Geography of Asia</td>
</tr>
<tr>
<td>497 North American languages</td>
<td>916 Geography of Africa</td>
</tr>
<tr>
<td>498 South American Languages</td>
<td>917 Geography of North America</td>
</tr>
<tr>
<td>554 Earth sciences of Europe</td>
<td>918 Geography of South America</td>
</tr>
<tr>
<td>555 Earth sciences of Asia</td>
<td></td>
</tr>
<tr>
<td>556 Earth sciences of Africa</td>
<td>940-949 History of Europe</td>
</tr>
<tr>
<td>557 Earth sciences of North America</td>
<td>950-959 History of Asia</td>
</tr>
</tbody>
</table>
However, unlike the biological case, terms keep recurring and it is arbitrary, which criteria one gives priority to. For example, in this case, an obvious alternative is to use geographic area before disciplinary area, which yields a completely different classification of books. Indeed, there are as many equally correct ways of classifying books as there are adjectives to use to describe books.

The contrast between a classification of convenience, such as that used for library books, and biological classifications, was apparent to naturalists before Darwin. It really seemed that there was only one meaningful way to classify organisms, but many equally reasonable ways to classify books: classification of books (and the like) is “artificial” that of organisms is “natural.” To the pre-Darwinian taxonomist, the only explanation of the naturalness of biological classification was that there was some preexisting structure in God’s mind that was represented in his creation.

One of Darwin’s great insights was that a nested and hierarchical classification can be explained without recourse to the “mind of the creator,” if the traits used to classify organisms had evolved along the branches of an evolutionary tree. For example, consider the vertebrate traits vertebral column, lungs, legs, amniotic egg, feathers, wings, fur, and milk. If we map these traits on the tree of vertebrates, as we now understand it, we can see that it matches the Venn diagram given earlier.
Given this history we can understand why a classification based on these traits forms a “natural” nested hierarchy. Why, for example, the amniotes (a taxon/clade defined by the amniotic egg) is divided into three non-intersecting groups: mammals (with fur and milk), birds (with wings and feathers) and reptiles with neither. Likewise, it is explicable why all animals with feathers would also have a vertebral column, four limbs, lungs, and an amniotic egg. If classifications group organisms based on their evolutionary kinship, what Darwin called “propinquity of descent,” then we expect traits to show a nested structure and to yield naturally hierarchical classifications. Thus the tidiness of biological classification can be explained most simply as a result of trait evolution along the branches of a phylogenetic tree.

The evolutionary explanation of classification not only accounts for cases of perfect hierarchical structuring of traits, but also explains cases that are less “tidy.” Whales possess many traits of Mammalia, including lungs, pelvic and pectoral girdles, an amniotic egg, hair (eye-lashes), and milk, but they lack the four limbs that are typical of Mammalia (and other Tetrapoda). A religious pre-Darwinian scholar would not explain this by invoking sloppiness on the part of the Creator and, therefore, was challenged to find some special purpose or message in such anomalous patterns. In contrast, the post-Darwinian naturalist could well understand that hind limbs were lost when whales invaded the oceans and adapted to an aquatic way of life [Figure].

*Monophyly and Phylogenetic Systematics*
As discussed in chapter 1, there was a surprising delay between the acceptance of evolution (mid-19th century) and the recognition that taxonomy needed to be changed to reflect evolution (mid-20th century). Following Darwin, taxonomists did become more careful to avoid lumping distantly related organisms based on convergently evolved traits. But until the development of Phylogenetic Systematics, classifications were still built around “important” traits, even at the expense of accurately reflecting evolutionary relatedness. For example, despite the recognition that the vertebrate phylogeny was similar to that shown, living vertebrates were traditionally divided into the five classes shown in the figure.

This classification seemed sound: each of the classes was readily distinguished from the others by one or a combination of traits. Only one trait, limbs, showed homoplasy (Chapter 3) and this is not problematic because all groups that have lost limbs have close relatives that have retained limbs. The data matrix below summarizes the features of the five classes.

<table>
<thead>
<tr>
<th></th>
<th>Vertebral column</th>
<th>Limbs &amp; lungs</th>
<th>Amniotic egg</th>
<th>Wings &amp; feathers</th>
<th>Hair &amp; milk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphibians</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reptiles</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birds</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mammals</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

However, while this classification is compatible with these traits, it fails to properly represent evolutionary relatedness. Specifically, the taxa Pisces and Reptilia have problems. A lungfish (in Pisces ‘d’) is more closely related to members of Amphibia, Reptilia, Mammalia, and Aves than to the other three living lineages of Pisces: ray-finned fish (Pisces ‘c’), sharks and rays (Pisces ‘b’), and lampreys (Pisces ‘a’). Likewise, a crocodile (in Reptilia ‘c’) is more closely related to a bird than to a turtle (Reptilia ‘a’) or
lizards and snakes (Reptilia ‘b’). The problem is that neither Pisces nor Reptilia correspond to clades: neither is a monophyletic taxon. You can see this because it is impossible to separate all the tips assigned to Pisces or Reptilia from the rest of the tree with a single cut, which shows that these groups do not include a single ancestor and all of its descendents (see Chapter 2).

As the pruning and collapsing maneuvers (Chapter 2) show, members of a clade occupy the same position in the tree of life: they are more closely related to each other than to any organisms outside the group. As shown in the figure, all members of clade A are more closely related to any other member of clade A than to any organism outside clade A.

Non-monophyletic groups, in contrast, occupy multiple distinct positions in the tree of life. Consequently, some component organisms are more closely related to organisms outside the group than to organisms inside the group. For example, the non-monophyletic group labeled B includes some organisms (in clade B1) that are more closely related to organisms in clade A than other organisms in group B (specifically, clade B2).

But how does a classification based on homologous traits yield non-monophyletic taxa? Part of the problem is that the classification is based on shared ancestral characters. The shared traits of Pisces are ancestral traits for all vertebrates: they are really just the lack of the traits of land vertebrates. As a result, the label “Pisces” basically refers to all vertebrates that are not tetrapods. So while Pisces has a diagnostic set of traits, Pisces is not a monophyletic taxon. Compare this to the birds, which are united by the possession of wings (among other traits). Wings evolved on a particular branch, and thus all descendants of that branch are expected to have wings (Chapter 3). Wings therefore demarcate a monophyletic group (clade) that corresponds to birds.

From this example, we can induce the general principle that monophyletic taxa (clades), and only monophyletic taxa, accurately reflect evolutionary relationships. This is why modern classifications assume that all taxa be monophyletic. More subtly, but equally importantly, traits have now shifted from being defining features of taxa to providing evidence of monophyly and facilitating the assignment of individual specimens to taxa.

When reassigning names that previously applied to non-monophyletic groups, membership in the taxon can be either expanded or contracted. These are illustrated well
by the fate of the names Reptilia and Pisces. Reptilia were contracted when fossils groups (e.g., pelycosaurs), formerly placed in Reptilia but actually more closely related to mammals, were removed. More significantly, Reptilia was also expanded to include all organisms more closely related to lizards and snakes than to mammals. Hence, the taxon Reptilia now includes the birds.

Rather than expanding Pisces to include all Vertebrates, this taxonomic name has fallen out of favor. Instead, modern taxonomies tend to recognize a number of smaller groups, each of which is monophyletic: the Cyclostomata (Pisces ‘a’: Hagfish and lampreys – although these may be two distinct lineages), the Chondrichthyes (Pisces ‘b’: sharks and rays), Actinopterygii (Pisces ‘c’: ray-finned fish), and the Coelacanthiformes (Pisces ‘d’: coelocanths and perhaps lungfish). The term “fish” now has lost all taxonomic meaning and instead refers to a body form and way of life.

**Varieties of non-monophyly**

It is common to distinguish two varieties of non-monophyly: Paraphyly and Polyphyly. This distinction is not very useful, but seeing as you will encounter them in scientific publications, a brief clarification is required.

Paraphyletic groups are ones that were formerly recognized as a taxon because of shared primitive characters. Fish and reptiles, in their old usage, are paraphyletic groups. In each case they are united by the possession of ancestral traits of a larger clade (Vertebrates and Amniotes, respectively) combined with the lack of derived traits (those of Tetrapods and Mammals plus birds, respectively). In a paraphyletic group if one were to look at the most recent ancestor of all members of the group it would have all the traits of the group.
Polyphyletic groups are ones that are based on convergently evolved characters. For example, the group “Apoda” was at one time composed of snakes and various legless lizard groups, but we now know that snakes and lizards lost limbs independently: leglessness is not homologous in all members of “Apoda.” In a polyphyletic group, if one were to look at the last ancestor of all members of the group it would not have the traits of the group: the last common ancestor of all “Apoda” had legs.

An alternative way to distinguish paraphyly and polyphyly is based on the identity of the last common ancestor of all members of the group. If the last common ancestor is in the group (and the group is not monophyletic), then the group is paraphyletic. If the last common ancestor is outside the group, then the group is polyphyletic. However, the question of whether the organisms living at an internal node of a tree are or are not members of the group has to be based on information on those organisms’ traits. Therefore, this criterion for distinguishing polyphyly and paraphyly is best viewed as a less direct application of the trait-based distinction presented above.

**Converting monophyletic classifications into trees**

Because a tree is composed of a set of non-overlapping clades, a monophyletic classification will be perfectly consistent with the tree upon which it is based. This means that one can represent a classification in tree form. Being able to interconvert textual or graphic representation of a classification into a tree and vice versa is an important tree thinking skill. Once developed it becomes possible to take information in a classification and connect it seamlessly to other kinds of evolutionary data.
To convert a tree into an exhaustive classification, you need only list the clades that need to be named and then arrange them in either a Venn or indented list format (the two most common ways to represent a classification). The Venn format is rather unwieldy in practice, but we think you will easily see how it mirrors the tree. The indented list format, however, may need more explanation. To make an indented list from a tree, visit each internal node on the tree and assign it a label, which will be taken as the temporary name of the clade descended from that node. Then, you list the clades with less-inclusive clades indented relative to the clades that include them.

1) First list the name of the most-inclusive clade: clade A
2) Indent by one tab, and then list the two subclades of A: clades B and C
3) Below clade B, indent, and then insert the names of its two subclades: Species 1 and 2.
4) Below clade C, indent, and then insert its two subclades: Species 3 and Clade D.
5) And so on.

Hopefully you can see that not only does this tree imply this (and only this) indented classification, but given the classification you could draw the implied tree.

**Taxonomic Ranks**

The first development of formal taxonomic ranks goes back to the Swedish botanist Carolus Linnaeus. He focused his attention on what he perceived to be the two basic ranks: genera (singular = genus) and species. He believed that God had created broader kinds, genera, which had subsequently diversified following the expulsion from Eden to form multiple manifestations, species. The genus rank is above the species rank: a single genus can contain multiple species, but a species is only assigned to one genus.

To aid in distinguishing the names of species and genera (and in line with his philosophy of kinds), Linnaeus proposed a *binomial* system of naming, in which a genus name is
capitalized and a species name is composed of both the genus name and the lowercase (adjectival) species name. Both names are conventionally italicized or underlined: the correct name for the human species is *Homo sapiens*, which may be abbreviated *H. sapiens* (once *Homo* is indicated), but is never just *sapiens*.

Over the 19th century a richly subdivided taxonomic hierarchy emerged, with an agreed upon set of ranks from the least inclusive (lowest) ranks, species (sometimes subdivided into varieties) and genus (sometimes subdivided into sections or series), to the most inclusive (highest) rank, kingdom. Between these extremes were a set of intermediate ranks including, from lowest to highest, tribes, families, orders, classes, and phyla (sing. phylum). And when these did not allow enough scope for organizing taxa, intermediate ranks could be indicated (e.g., superfamily, subgenus, infraorder).

It is an interesting fact that biologists widely use ranks but generally agree that ranks have little inherent meaning. While we might imagine that higher ranked taxa are generally older, larger, and more diverse than younger ranked taxa, this is not always the case. You cannot assume that a tribe of palms is younger than a family of beetles, nor that it less distinct from its close relatives, nor that it contains fewer species. It would, thus be potentially misleading to use ranks to compare groups. To conclude that Brassicaceae contains fewer species or less phenotypic diversity than Malvaceae because it includes fewer subfamilies would be a dangerous assumption. So what, if any, role do taxonomic ranks play?

The only information conveyed by rank is the nesting of clades: higher ranked taxa contain lower ranked taxa, not the reverse. So, a clade at one rank cannot include a subclade that is at an equal or higher rank. It is also generally argued that sister clades should be at the same rank. These principles constrain the ranks of clades, but it is always possible to assign all clades to ranks in such a way that this rule if followed. The figure illustrates this with an example. You will also see that in order for this pattern to be sustained, some taxa are *monotypic*, they contain only one of the next lower rank. For example, genus 1, is the sole genus in tribe 1, which is the sole tribe in family 1.
The figure shows that ranks can be applied to monophyletic taxa, but that doesn’t answer the question of whether ranks serve any useful purpose. In fact, the primary role of ranks is not to communicate biological information but to provide an anchor for the names of groups. Therefore, to understand ranks, and why they are currently controversial, we need to briefly examine taxonomic nomenclature: the systems by which taxa are named.

**Nomenclature**

The most basic role of taxonomy is to provide a stable tool for scientific communication, by attaching names to taxa. When somebody talks about “Lepidoptera,” for example, it is important that this term have an unambiguous meaning. Otherwise, we will too often find ourselves talking at cross-purposes. The system by which scientific names are attached to taxa is called nomenclature.

In the 19th century as the number of practicing taxonomists increased, large numbers of new taxa were named and many traditionally recognized taxa were redefined. This resulted in great instability and confusion in taxonomic circles. As a result, zoologists, botanists (which then included people studying fungi, algae, and protists), and bacteriologists each developed formal rules to regulate taxonomic naming. The resulting codes of nomenclature were developed long before phylogenetic principles (or even evolutionary ones) were widely accepted. One manifestation of this disconnect is that the these codes heavily emphasize ranks, even though the evolutionary meaning of a rank is unclear. The centrality of ranks is best shown by three features of the traditional codes. First, they state that all organisms must be assigned to a few mandatory ranks: e.g., species, genus, and family. Second, standardized endings were proposed for many ranks (table X), meaning that the name of group changes if its rank changes. Third, and most importantly, the correct name of a taxon is determined by the principle of priority at rank. This states that the correct name for a taxon (defined by reference to particular specimens called types) is the earliest name published for the group at whichever rank it is to be recognized. This means that you cannot determine the name of a taxon unless you know its rank.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Botanical ending</th>
<th>Zoological ending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tribe</td>
<td>-eae</td>
<td>-ini</td>
</tr>
<tr>
<td>Subfamily</td>
<td>-oideae</td>
<td>-inae</td>
</tr>
<tr>
<td>Family</td>
<td>-aceae</td>
<td>-idae</td>
</tr>
<tr>
<td>Order</td>
<td>-ales</td>
<td>Not specified</td>
</tr>
<tr>
<td>Class</td>
<td>-opsida</td>
<td>Not specified</td>
</tr>
<tr>
<td>Phylum</td>
<td>-phyta</td>
<td>Not specified</td>
</tr>
</tbody>
</table>

With the development of phylogenetic systematics, taxonomists acquired an objective basis for deciding if a group was a taxon – only monophyletic taxa being considered valid. However the naming of clades depends on ranks, which are ill-defined: nobody can say for sure whether a clade is a family or an order. It was natural, therefore, that there would be an effort to think clearly about how names are attached to groups and examine the central role that ranks play in the process. A new system of nomenclature, phylogenetic nomenclature, was developed and gave rise to a new code of nomenclature,
the PhyloCode. This is still controversial and may never become widely adopted, but is nonetheless worth briefly summarizing.

The central idea of phylogenetic nomenclature is that one can attach names to clades using one of a few methods. Even if you will never be engaged in nomenclature per se, it is worth knowing how names can be attached to clades, because this provides a useful way to communicate about trees. The two most important clade-naming methods are: node-based and branch-based clade definitions.

A node-based taxon is defined as the least inclusive clade that includes a set of specifiers that are listed. For example, I might choose to associate the name Mammalia with the least inclusive clade that include a duck-billed platypus, an opossum, and a human (each of these specifiers is supposed to be attached to a particular specimen in a museum). As you can see in the figure once we have a tree and know where the specifiers fit, we can immediately determine the clade to which the name Mammalia should apply. Mammalia will comprise organisms that are all more closely related to platypus, opossum, and human than to any other organism.

A branch-based taxon is defined as the most inclusive clade that includes some listed internal specifiers but excludes some listed external specifiers. For example, Hominoidea might be defined as the most inclusive clade that includes gibbons and humans but not baboons. Again, the correct application of a branch-based name is dictated only by the phylogenetic resolution. In this example, Hominoidea includes all organisms that are more closely related to humans (and gibbons) than to baboons.
As you can probably see it ought to be possible to generate a database of phylogenetic names attached to a current best estimate to the tree of life and thereby get rid of a lot of confusion as to what scientific names refer to. Unfortunately, such a system is not in effect, but we may hope that something like it emerges over the next decade or two.

**Major Points**

The guiding principle of biological classification is to assemble groups of organisms that are more closely related to each other than to any organism outside the group. Because relatedness is based on common ancestry, monophyletic groups form the basis of classifications. The hierarchical structure of classification is not arbitrary but explicable due to the tree like nature of evolutionary history.