

An Introduction to Evolutionary Biology

Prof. Sutirth Dey

Biology Department, Population Biology Lab

Indian Institute of Science Education and Research (IISER) Pune

Week 2 Lecture 8

Evidence for evolution 2: Descent with modification Part 1

So, in our last discussion, we started talking about the evidence for evolution. And we said that this essentially boils down to three different kinds of questions. The first question is, do species really change? We saw with multiple examples that the answer is yes. Then the question was whether species can arise from existing species, and the answer was again yes. And then the third question was, given all this, Can we really say that all the biodiversity around us has actually arisen due to descent with modification? And that is what we are going to consider now. Now, before we get to this, the major question is that What we need to ask is, what exactly do we mean by the term "descent with modification"? So, descent with modification refers to the notion that species change over time which finally leads to the formation of new species. And what exactly are their implications? This is what we have in order to understand its importance; that is what we are going to look at. So, let us assume that you know that in whatever we are going to discuss, time flows from top to bottom. And let us say we are starting with an ancestral group, which has faces.

So, at some point, due to descent with modification, this divides into two groups. One of them basically retains ancestral properties and continues to have faces right up until modern times. The other one, while forming, somehow acquires a mutation. Or a bunch of mutations due to which, let us say, it starts now having eyes. Now, this species, this group, at some point, again splits into two.

One group has, let us say, ears on top of the eyes, and then that group does not change any further. And it essentially comes down to modern times with ears and eyes. The other group, let us say, acquires another mutation or another set of mutations by which it now also acquires a mouth. And let us say the process continues, and at some point, this one, which has a mouth, splits into two. One of them essentially just continues the ancestral state of having a mouth whereas, the other one now let us say adds a tongue to this. And let us assume you know, just for the sake of discussion, that this group now stays as it is.

So, if you look at the present times, you have four species that look somewhat like this. Now, remember this information I do not have, whereas this information I have includes these four kinds of species. Now, if I look at these four species, this automatically leads to what is known as a nested hierarchical organization.

What do I mean by that? So, if you look at it, you have, let us say, this species right at the, you know, on the right-hand side. Which has a face, has an eye, mouth, and tongue. All organisms have this; therefore, we can recognize it as a species. Now, you have another nearby species, let us say, which has everything else but does not really have the tongue. So, now you can take this species and that species and group them together.

Because both of them have the common properties of the face, the eyes, and the mouth. So, if you call this innermost group the species, you can probably now call this group. And this group, together as a larger group, says something like the genus. Now, you can extend the process further, and you can compare these two. and then also include this one, which is different from these two in not having a mouth.

But you can take all three of these as a group of organisms that only have a face and eyes in common. And you can call this a higher order; let us say you call this whatever you call this, let us say a family, right? And similarly, now you can take all four of them together and make them into some hierarchical group of organization. which has only face as the common property across all four of them And let us say you call these things whatever class, let us say. So, you can see where this is leading; this is exactly what Carolus Linnaeus was trying to do. When he was coming up with his scheme of classification.

And please appreciate that Carolus Linnaeus was not the first person. To do this kind of nested hierarchical classification of living organisms. His was, of course, the most well-known, the most thorough, and comprehensive. But this kind of nested classification is much, much older. I mean, even Aristotle is known to have done some kind of classification like this.

So, descent with modification explains very nicely why this kind of taxonomic classification is possible to begin with. And, also appreciate that this kind of classification can be done for multiple groups. I mean, you can do it for animals; you can do it for plants; you can do it for fungi, etcetera, etcetera. So descent with modification trivially explains why one sees the ability to do such a thing. However, we are going to have a problem here.

Remember, we are trying to see if there is evidence of descent with modification. So, the

fact that this kind of nested hierarchical classification is possible, Does that necessarily imply that all these organisms are evolutionarily related to each other? So, this is like that logical fallacy that we were talking about: affirming the consequent, okay? If A will lead to B, and B has happened, does that mean that A has caused it? So, just to put it formally, just because one can produce a scheme to arrange things does not guarantee that. The things that have been arranged have some kind of relationship with each other in terms of how they were produced. So just to give you an example, suppose you go to a library, and you see that the books are very nicely arranged. Typically, in any library, the books are arranged by topics, right? So, you have, let us say, all the science books on one side and all the social science books on another side.

All the literature books are on the other side. In the science sections, you find, let us say, the life science books arranged together. The physical science books are clubbed together, and the chemical science books are clubbed together. Then within the biological sciences, you have, let us say, biochemistry books together. Evolution books, ecology books, and so on and so forth.

Now, this is exactly the kind of nested hierarchical scheme that we are talking about in the context of linear syntaxonomy. But in the context of the books, just because you have this kind of nested hierarchical classification, does not really tell you that one book arose from the other, right? In fact, you know that they did not arise that way; most of the books actually arose independently of each other, correct? Therefore, the fact that such classification schemes are possible by itself does not guarantee that. Descent with modification has occurred. So how will you solve this problem? What happens if we look at the same object, but we try to classify it using some other properties? So remember when we were classifying those emojis? We classified them based on the presence of, let us say, You know, ears, eyes, and so on and so forth. Can you do it using something different? So, as if you have the organisms that you are classifying, if they have not arisen through descent with modification, Then you do not really expect that using different schemes of classification will lead to similar results.

On the other hand, if they have indeed arisen through DWM and if they are related to each other, Then we expect the results to be at least roughly similar. So, just to go back to our library analogy again, suppose instead of organizing the books by, let us say, subject, Let us say now you organize them alphabetically by, say, the author's name. So, all the books whose author's name starts with A are grouped together. All the books whose authors' names start with B are grouped together, and so on and so forth. Whatever arrangement you now get, Do you really expect that to be the same as the arrangement that you get when you arrange them by subject? Obviously not, right? That is because, as we pointed out, the books are actually not related to each other.

They were not, you know, one book did not give rise to another book. However, in our context, if we end up using, you know, morphological features, If we end up using some other information to figure out these evolutionary relationships, you know. How they are related to each other, or even taxonomically, you know how they relate to each other. Then what is going to be the result? But before we ask that question, we have to ask what other information we can use. And this is where molecular taxonomy enters.

So, if you remember when we were dealing with the history of the subject, We said that as soon as people figured out how to sequence proteins and how to sequence genes, One of the first things they did was try to sequence proteins and genes from other species. and then trying to figure out what the relationships are between those organisms based on those sequences. And just to cut a long story short, it turns out that there is a very good degree of overlap. Between the results of traditional and molecular classification schemes, particularly at higher taxonomy levels. So, at the level of the class and the family, the relationship that you get.

Based on morphological versus molecular methods, they are very close to each other. The evolutionary trees look very close to each other. The lower you go in the sense of the level of genus and species, and so on and so forth. That is where you end up classifying one species from one genus to another genus, and so on and so forth. Now, if you think about it, this is not very unexpected; why? Because the number of morphological traits.

The amount of information that is available for you to do the classification is actually much, much less compared to the amount of molecular information. that is available to do the things you know to figure out those relationships. Therefore, the degree of certainty and the degree of resolution that you will have from any molecular scheme is important. is far greater than the degree of resolution that you will have from a morphological scheme. Hence, a better way of comparison is actually to compare trees that have been made based on one molecule.

One kind of molecule with, let us say, another kind of molecule. So, let us say protein A compares it with protein B or gene A compares it with gene B. And again, just to let you know, give you the spoiler straight away: even when you do it this way, it turns out that. The relationships, also technically known as phylogenetic trees, that you construct using multiple genes. Or multiple proteins, more often than not, actually lead to extremely similar results.

However, before I give you examples of that and show you what I mean by very similar, what degree of similarity is very similar? I first need to take a quick digression, and we

need to discuss what exactly a phylogenetic tree is. I am sure you have heard of this term now and then and how one can construct it using particularly molecular sequence data. It is not obvious how one can do it. So, just to ask, what exactly is a phylogenetic tree? A phylogenetic tree is a graphical representation of the evolutionary relationships between biological entities. And when I say entities, you can express it in terms of relationships between species or relationships between proteins or a relationship between genes, whatever. So, to give you an example, let us say we look at humans and we look at the two close relatives, chimpanzees and bonobos. And we ask the question of how exactly these species are related to each other. So, this is the form that the relationship takes. In a moment, we will discuss how we get to this kind of form.

But for a moment, let us assume that this is the nature of the relationship. So, what we want to show you is how exactly one reaches here. So, in this kind of diagram, you know time runs from one end to the other; it can be from top to bottom. Bottom to top, right to left, whatever, but it will always be shown in which direction it is running. So, for example, in this particular case, time is running from bottom to top.

So, on one side, you know whatever is the present side of time, You are going to have the three existing groups, which in this case are humans, chimpanzees, and bonobos: H, C, and B. Now, these three are the existing species, and therefore, they have not yet led to something else. which is why they only have their ancestors connected to them, but they do not have anything else to which they are connected, Which basically means that these are at the terminal ends of the tree, and that is why they are known as the terminal nodes. Now, if what I am showing you over here is what you know is the ancestor from which These two guys, the chimpanzees and the bonobos, have arisen. So, if you look at it, this is the one that is, you know, marked with a D.

So, this graph is showing that, you know, D is the common ancestor for both chimpanzees and bonobos. and this line, which is connecting this common ancestor to the existing species, This line is what is known as the lineage or branch, you know, in relation to a tree. Now, this relationship is also showing that this common ancestor of chimpanzees and bonobos. This itself arises from another ancestor known as G. And this G is what is known as the most recent common ancestor for this particular group.

So, for all three species that I have, this G is the most recent common ancestor, and this green thing over here. This is what is known as a root; this is basically depicting the point that a bunch of, you know. Species have a bunch of ancestral lineages they have come from, and they have led to the arising of this G. So, we are not interested in what has happened prior to G.

But we are interested in G. Because this is the most recent common ancestor of all three species that we are interested in. So, this is showing that this tree indicates two different species and two different lineages arose from G. One is called E; the other is called D. From E arose two groups; one was called F, but F went extinct in this particular case. And the other group is what finally led to humans.

On the other hand, as I said, you know that G arose from D, and then D led to C and B, chimpanzees and bonobos. So, the crucial information that you get out of this kind of tree is what the amount of relatedness is between the various species. And the simple way to think about it is that two species are more closely related if they have a more recent common ancestor. So, if you look at it, you know, remember we have time on this axis and time on the y-axis. So, chimpanzees and bonobos have a common ancestor in the form of D, which is roughly at this point whereas humans and chimpanzees share a common ancestor, this G, most recent common ancestor which is much earlier in time. Therefore, what we will say is that chimpanzees and bonobos are more closely related to each other, compared to humans and chimpanzees or humans and bonobos, right? And you can see that it is happening because this common ancestor for chimpanzees and bonobos is later in time compared to the other one. Now, this is a very simple way of thinking about a tree, you know.

In this particular case, we are just saying time, but we are not really calibrating it. We are not saying how much time has elapsed between this and that. There are certain kinds of trees in which the length of the branch is calibrated to the amount of time that has passed. In other words, the longer the distance between two nodes, the more time has elapsed between the two nodes and vice versa. So, this is not that kind of tree; this is a much simpler tree where we are just showing the relationship and nothing else.

So, now that you know these terminologies, the question is, how exactly does one get these kinds of trees, particularly from sequence data? Once you understand how it happens from the sequence data, you will realize that how it comes from morphological data is also very similar. But let me particularly start by giving you an example using a sequence of data. So, let us assume that we have three species; we will just call them two, three, and four—these three.

And let us say we are interested in understanding the evolutionary relationship between the three of them. So, basically, we want to know who came from whom and, roughly speaking, who diverged earlier and who diverged later. Now these three groups in which we are interested are what are known as the in-group. Now, in order to do this kind of, you know, reconstruction of the phylogeny, we require one group that we definitely know is not part of the in-group.

We are not interested in it, and we definitely know that this is something that diverged much, much earlier than the in-group. And in this particular case, that happens to be this species 1, and this is the one that we are going to call an outgroup. These are technical terms. So, the whole idea is that we start with the assumption that the outgroup's sequence is the ancestral form. And essentially, the question we are asking is this: How does one go from the outgroup's sequence to the various ingroup sequences? In other words, what kind of mutations need to happen? What kind of changes need to happen to go from 1 to the existing patterns of 2, 3, and 4? How does one do that? So let us see.

So, remember we said that the outgroup sequence is the most ancient one. So, we take that right over here, and therefore, We assume that the outgroup, right, you know, splits right at the beginning, and it has not changed ever since. So, this is my outgroup. Then we assume that this outgroup split, or rather, this common ancestor of the outgroup. And our ingroups split at some point; this is over here, and then from here, this split into two sections.

One lineage, or two lineages—actually, one lineage—had this mutation, which converted the A at the ninth position to a T. Remember that this is the only difference between 1 and 2: this A at the 9th position has become T. So, that is this mutation, and let us assume that the other lineage acquired another mutation at some point. And after that mutation, this common ancestor, you know, split into two groups: one led to three, and the other led to four. Why are we assuming this? Because remember, 3 and 4 actually have a more or less identical sequence and the main thing is that you know both of them; they have this A over here. which is different from this T, and they have this A over here, which is the same as this one. So, basically, they differ from this at exactly one locus. So, we end up, yeah, so if we end up thinking of it like this. Then we can see that just two changes are enough for me to get all the, you know, sequence information that we have.

So, we just need to assume that one change has occurred in this lineage. Another change has happened in this lineage at this point before this common ancestor, and that is it; we will get this back. Now, obviously, this is not the only way to get to this result; there are other ways too. But in these other forms, you will see that. You actually need to posit a much greater number of changes in order to reach the same situation.

So, for example, suppose you know, let us look at tree number 2. So, in tree number 2, we can assume that you know about this other outgroup split. And after that, there is a common ancestor for 2, 3, and 4, and for this common ancestor for 2, 3, and 4. One lineage had this mutation, this third position T becoming A, and in the other lineage again. There was one split over here, one common ancestor, and this lineage over here

has this mutation, this ninth position mutation. and the other one has this mutation; the other one has again the same third position A to T mutation sorry, T goes to A mutation.

If you do that, you will essentially recover the same tree. But when you are doing this, remember that you have to posit at least three mutations to get to this. The same situation over here; just, you know, 3 and 4 have interchanged places. So, again, you need to posit that you require at least three mutations. So, what we are seeing is that there is one tree in which I can do it with just two mutations.

But there are two trees in which I can do it with three mutations. Of course, there are other trees possible that will lead to the same results with many, many more mutations. But there is no other tree that will lead to this same result with fewer than 2 mutations. So, what we say is that this first tree leads to a relationship with the least number of mutations. That is the most parsimonious tree; it is the tree which is the. which requires us to make the minimum number of assumptions about mutations and that is why we assume that this first tree is the correct relationship between these three species: 2, 3, and 4. Now, this is just one way to make a tree. I am not saying that this is the way to make a tree. There are many other ways, and you know, given your knowledge of various kinds of biological realities, Sometimes the other ways of making the trees are more powerful. However, this way of making a tree is actually the simplest for us to understand conceptually and incidentally, this is more or less what is done even in the context of, you know, characters. So, you know when you are looking at morphological trees? Most people use this parsimony method to figure out what the evolutionary relationship is. So, for example, if you take all the properties of, you know, all the vertebrates and you try to classify them. You know, in terms of their evolutionary relationship, you will see that you know. At some point, the vertebrae arose; obviously, that is how you have vertebrates and then you know those are the sharks, and then at some point you know. The skeleton becomes bony, and you know from that that after that, you get the ray-finned fishes. Then at some point, we develop four limbs; that is where you have tetrapods. So, these are the last tetrapod ancestors, and then these tetrapod ancestors, you know. Diverging to become the amphibians, one group of tetrapods ended up having what are known as amniotic eggs. So, if you look at the eggs of fishes and amphibians, they have to be laid in water.

Because they are extremely sensitive to desiccation and cannot be laid in normal air, They will simply end up, you know, drying up. So, at some point, the tetrapods ended up coming up with, you know, basically watery eggs. So, if you remember the egg of a chicken, you will see that there is an outer hard shell and there is a membrane. known as the amniotic membrane, which essentially ensures that the water inside is not, you know, dried up or desiccated. And that is why the egg, the developing embryo, is actually able to develop in essentially a watery environment.

So, that is what is known as an amniotic egg. So, then you have this ancestor with the amniotic egg common ancestor, and one group ends up, you know. Developing hair on their bodies, this group then split, and one lineage became primates. another lineage goes and becomes rodents and rabbits. On the other hand, the common ancestor with amniotic eggs developed at some point. What are known as, you know, two post-orbital fenestrae? So, basically, this, you know, is the orbit of our eye, and behind this orbit, there are holes; these are known as fenestra and this group of organisms leads to the dinosaurs and birds on one hand and to crocodiles on the other, These are the organisms that actually have two such holes, which allow, among other things, for jaw muscles to attach. So, this is the evolutionary relationship between these various organisms, and this can essentially be made. Based on precisely the kind of logic, parsimonious logic, that we applied to the sequence case. Anyway, this was all about how one makes trees, particularly using sequences.

But remember, our primary question is that if we have trees that have been made using one method, let us say using morphological data. and trees that have been made using other methods. Then how well are they, you know, corresponding with each other? We said that they are correspondingly fantastic. And then we said that even trees that are made with different genes or different proteins, even though they correspond very well to each other.

Now that is the one for which I want to give you an example. So, in 1982, Penny et al. took 5 different proteins from 11 different species of mammals and essentially sequenced them. Now they built trees based on phylogenetic trees using the same logic that I talked about. Given the number of species and the number of proteins, The total number of possible trees is 3 crores, 44 lakhs, and 59,425, which basically means more than 34 million. Now, if these proteins are not, you know, if these mammals have not descended with modification, Then we do not really expect them to be related to one another. and then we do not really expect the trees to be very similar to one another, Particularly given that more than 34 million of them are possible.

However, if they are related by DWM, then the across-species relationships shown by these different proteins should actually be extremely similar, which is exactly what we get. When we look at the most parsimonious trees for each of the 5 proteins, We find that they are remarkably similar to each other, and the major relationships between, you know, Which group, which mammals are clustering with each other and which are clustering away from each other? That major relationship is actually remarkably conserved. So, as I said, if the 11 species had originated independently of each other, Then the probability of the trees constructed out of 5 different proteins being similar would actually have been infinitesimally small. The fact that they were similar shows unequivocally that they came

about due to descent with modification.

And again, remember this is just for 11 mammals; you can do it for many other organisms. You keep getting back this observation that when you do it with multiple traits. Or when you do it with multiple protein sequences or gene sequences, you are getting back very similar trees. And this fact that it is generally true across all living organisms is what makes descent with modification significant. Such strong evidence is what makes phylogenetic trees such strong evidence for descent with modification.

To quote Charles Darwin himself, this is a long quotation. So, this is, you know, one of those old-style sentences where one sentence is half a paragraph, kind of thing. So, I have kind of broken it into multiple pieces so that we can phrase the sentence better. So, Darwin says that all the difficulties in classification have been explained. So, wherever you have those dot dots, those are the places where.

I have just edited out certain words simply because, you know, this is making the flow better. So, all the difficulties in classification are explained from the view that the natural system is founded on descent with modification. So, difficulties in classification are explained if you assume DWM is correct in that the characters. Those which naturalists consider as showing true affinity between any two or more species are. which has been inherited from a common parent.

So, all the most important characters have come from common parents and that community of descent. So, basically, different groups are descending with modification in different lineages; that is the hidden bond. Which naturalists have been unconsciously seeking. So, this is a statement that shows the importance of descent with modification in understanding. How different lineages have originated.

However, taxonomy and phylogeny are not the only evidence in favor of descent with modification. There are actually many more, and that is what we are going to look at in our next discussion. Thank you.