7.014 Introductory Biology, Spring 2005
Transcript – Lecture 31

So, for today's lecture as you can see up there is molecular -- evolution, and ecology. And what I mean by this, it's basically the study or what we try to figure out in molecular evolution and ecology is what genes or gene sequences can tell us about the evolution and ultimately also the ecology of organisms in the environment. And it's particularly relevant for thinking about microorganisms, prokaryotes and the environment. And I hope I can actually convince you today of that. This is interesting. The topics that I want to cover today is, first of all, I want to review a little bit what we know about life on Earth, sort of give an overview of the evolution of life on Earth. Then, I want to go into specific topic that's of particular relevance for the evolution of eukaryotes.

That's the endosymbiosis theory. And then I'll explain how we can use gene sequences to actually reconstruct events that have happened a very, very long time ago. OK, so we'll look at what we call molecular phylogenies, with the use of gene sequences to reconstruct the evolutionary history of organisms on Earth. Derived from that, we'll look at what we call the tree of life. That's sort of the big picture overview of the evolutionary relationships of all organisms on the planet. And then finally, I'll introduce you to a topic called molecular ecology. Again, that's how we can use gene sequences to learn something about the diversity of microorganisms in the environment that lead us then, next time, when I come back on Monday, into this big topic of environmental genomics, how we can actually expand this analysis to learn much more about organisms in the environment.

So, first of all, let's look at life on Earth. Does anybody know how old we think Earth is? Say again? Yeah, 4.5 to 4.6, I haven't my notes 4.6. So, Earth's thought to have originated about 4.6 billion years ago. When did the first solid rocks appear on earth? So, when was the surface kind of solidified? Anybody know? About 3.

9 billion years ago, OK? And when do we think life started to develop on the planet? Any ideas? Take a guess. Two? One? 3.5 billion years ago, OK? So, this is really remarkable. We think it didn't, I mean, of course it took a long time because were talking about millions of years and hundreds of millions of years, but still, if you look at the big picture, it didn't actually take life that long to evolve on the planet. So, why do we think that is the case? What's the evidence for that? Well, we look into sedimentary rocks, so old rocks that arose from sediments, what you find around this time, you find that chemicals start to appear, organic molecules that really resemble organic molecules in modern life. So, we have sort of chemical tracers, or chemical fossils.

So, tracers that indicate the presence of organisms. But what we also find is so-called micro-fossils, and I have a picture of that here where when you actually take rocks and actually slice them into very, very then slices, you can put them under specific microscopes. And what you then find is that many rocks that are very, very old, have those kinds of inclusions in them. And these things really resemble very much modern prokaryotic cells, modern bacterial cells, for example. And so, those micro-fossils are generally taken as an indication, also, that life is already present

during those times. Now, when we take a quick sort of overlook of the evolution of life on the planet, again this graph here summarizes sort of the last 4.

6 billion years or so when life originated. We see that there was a period of chemical evolution, and then somewhere here that region, it's, of course, not really well understood when that exactly happens, the origin of life is placed. But I want to alert you to a couple of really, really critical steps here that are shown on this graph which we'll actually talk more about. It is thought that life very early on is split into three major lineages: the bacteria, the archaea, in what is called here nuclear line. And I'll come back to that in a minute or so. Then, a further major event which you may remember is oxygenic photosynthesis actually evolved -- -- which means that cyanobacteria evolved that started to produce oxygen as a byproduct of photosynthesis.

And that really fundamentally changed the chemistry of the Earth. It actually became an oxidizing atmosphere. And what you see here is, once the oxygen concentration goes over a certain level, it allowed the development of an ozone shield. Now, what does that mean? What was the critical significance of the presence of an ozone shield? Does anybody know? What does it block out? Anybody remember that? What's the big significance of the ozone hole over Antarctica for example? It allows UV radiation to heat the Earth's surface, and in fact if there were no ozone, the UV radiation would be so strong that there would be no life possible on land. So, once the ozone shield actually developed, organisms could conquer, basically, the land's surface and settle on the land surface.

In this, then, is thought to be at least correlated with the development of endosymbiosis. And I'll explain what I mean by that. But it basically led to the origin of modern eukaryotes, so your ancestors essentially. But there was still a long time, obviously, until humans appeared. We have here the origin of animals and metazoans, and then the age of the dinosaurs is already a very small blip here on this graph.

And humans don't even get featured on that because we are so recent. So, but what I want to show you here is that three major lineages evolved early on. These are the bacteria, archaea, and what we call a nuclear lineage. And the significance of those nuclear lineages is that it basically combined with bacteria to form the modern eukaryotic cell. So, the eukarya, or eukaryotes they're also called. And it was this combination that we called the endosymbiosis event.

I want to explain this a little bit more, and then I'll show you finally why we actually know that those things are very likely to have occurred a long time ago. Yes? It means the bacteria and the nuclear lineages combine to form a eukaryote, OK? And I'm actually going to explain this on the slide here. So, if you have any more questions after that, please let me know. So, again, this shows you this early evolution, this early split in two archaea, bacteria, and this sort of nuclear line.

It is thought that this nuclear line, this was single celled organisms that increased in cell size, and then developed or partitioned the DNA into a nucleus, basically. So exactly how you find it in modern eukaryotic cells. But then what happened is the cell took up a bacterial cell, and over time this bacterial cell became symbiont. In fact it became the mitochondria. And so what this mitochondria now does in the moderate eukaryotic cell as you all know is it really took over the energy metabolism. So, the proto-eukaryotic cell took up a heterotrophic bacteria that form

the mitochondria. And this ultimately then gave rise to protozoa and to modern-day animals.

But there was a secondary symbiotic event. This cell, once it had taken up a heterotrophic bacterium, it took up an autotrophic bacterium, a cyanobacterium, an oxygenic photosynthesizer. And this actually that led to the development of modern algae and modern plants. So what we can say is that mitochondria our ancient heterotrophic bacteria -- And the chloroplasts are ancient cyanobacteria, so, oxygenic, photosynthetic bacteria.

And these obviously have coevolved to then form animals and finally your plants. So now, obviously we are talking here about events that happened a very, very long time ago. And so, the big question is really how do we really know this? But this takes me to the third topic, which is that of molecular evolution. So, we can state the problem again, And that is very simply put, evolution is incredibly slow, OK? And therefore, its processes are not directly observable.

And we need to actually use inference techniques to reconstruct evolutionary processes. Now, what do we use when we want to reconstruct the evolutionary history of animals and plants usually? Anybody? Fossils. Exactly. So you take a shovel, essentially, and dig down into the different layers. And there's different techniques that you can actually determine the age of different sedentary rocks. For example, and then you can construct, if you're lucky, you'll find enough fossils of a particular lineage.

You can reconstruct the evolution of the lineage. I'm sure you all have seen the example of the horse, for example, where we have actually quite good evidence when ancient horses look like. And we can reconstruct the sequence of events that led to the evolution of modern-day horses. Now, you can imagine, though, that when we talk about such ancient events like these there really is no fossil record.

OK, so what people have figured out, then, is that that was really a stroke of genius that came about in the late 60s, that DNA molecules can act as evolutionary chronometers. OK, now what do I mean by that? I mean that you can take DNA sequences or gene sequences from different kinds of organisms. Based on those gene sequences you can reconstruct the relationships to each other. You can determine whether two organisms are closely related or whether they are only very distantly related. And the underlying mechanism of that, is that mutations happen with a certain probability all the time.

So, the idea is that as time passed on, DNA molecules will change. So they will accumulate, actually, mutations, and so this will lead to, and that the idea is that the amount of change in a particular DNA sequence is proportional to the time of separate evolution of two different lineages or two different organisms. So, the amount is more or less proportional -- -- to time since the last common ancestry.

So, let me explain how this is actually done. What you really need in order to do this, is you need genes that are related to each other, OK? So, genes, they need to be universally distributed. That meets all organisms that you want to compare need to have this type of gene. And, those genes need to have conserved function. In these genes, we can then compare to each other, and I will explain how this is actually done. Any questions so far? OK, so the example that I actually want to bring is the 16S ribosomal RNA genes.

We oftentimes abbreviate this rRNA. Now, does anybody remember what the ribosomal RNAs are and do? What's the ribosome? Yes? Right, and what does it do? Exactly, it's the location where messenger RNA is translated into protein. Now, the ribosomal RNAs are an integral part of the ribosome. They play both a catalytic role as well as a structural role in the ribosome. And so, fundamentally, because this is such a fundamental organelle, all living organisms possess it. So, all organisms have it. So this allows us to use these genes to really compare all living organisms to each other. OK, so this is a very important point. I wanted to show you a, OK, if it wakes up.

There we go. An example of these ribosomal RNA genes, now this is actually, what you see here is a secondary structure of the actual RNA, the ribosomal RNA. Now, these molecules have a secondary structure because they play a catalytic and structural role. And so, the really amazing thing is when you look at the structure, the structure determines really the function of those molecules in different organisms. And then look at this.

We have here a bacterium, and here are an archaea. Now, if you think back to the first couple of slides, what I showed you is that those organisms have not shared a common evolutionary history for about four, or so, billion years, or 3 billion years, excuse me. But, if you just glance very quickly at the structures, you see that they look very similar to each other. So, there's an indication that the function is really very highly conserved of those molecules.

However, when you actually look at the sequences in detail, what you'll find is that there's different regions. And I'd given some examples here denoted by A, B, C in those molecules. And these different regions of the molecules are really the key to its usefulness in figuring out the evolution and ecology of many organisms. The region number A here, or denoted by A, a sequence stretches that are the same in all living organisms.

So they are universally conserved, which means that if you get a mutation in a gene in that particular region, you are dead. OK, that's why it's conserved essentially. Then we have those regions B where the length is conserved, but the sequence is not. So, there are sequence change allowed, but the length needs to be conserved. And then there's the region C were neither length nor sequence is actually conserved, and where we get a lot of variation. So, let me write this down. We have three types of sequence stretches. We have A, what I called the universally conserved sequences.

We have B where length, but not sequence is conserved. And, we have C where neither length nor sequence is actually conserved. And the first two stretches, the first two types of sequence stretches, are very important in figuring out the phylogeny or the evolutionary relationships amongst organisms. Whereas the sequence stretches number C because they vary so dramatically, are very important in identifying organisms. And we'll talk more about this actually next time.

So what can we actually know do with those sequences? Well, the first step is we need to generate an alignment. OK, and this is actually shown here, where each row denotes a gene from a particular organism. OK, so these are all abbreviated here. These actually aren't ribosomal RNA genes, but other genes. And that what you will see here is we can recognize those three different regions that I've pointed out

before. You have the regions A which tell you which nucleotides line up with each other, so you use this sort of as an anchor because the sequences never vary amongst organisms. And that the sequence region B where you light up sequences that vary or stretches that vary in sequence but not in length. Now, why is this important? It's important because you have in each column that nucleotides that have originated from a common ancestral nucleotide, and whose variation over time you can actually monitor.

Is everybody with that? Any questions? OK, great. The second step, then, is the calculation of a similarity. And this is shown here. Again, we have a very simplified alignment now of four different organisms. Here, we have the sequences that we want to compare. And what you'll see is that they're overall very similar, but there are different sort of nucleotides. And so, what we simply do is for each pair of sequence combinations, we calculate the sequence similarity value.

So, what you see is that you have 12 nucleotides, and the first pair differs in three nucleotides. OK, so that tells us, or it's called actually a distance here, I'm sorry. Let me write this down here. It's simply one minus the similarity, of course, but so basically a quarter of the nucleotides differ where it's between A and C, a third of the nucleotides difference on. OK, so you do this for each pair of sequences, excuse me. The third step, then, is to calculate the correction for multiple mutations affecting the same nucleotides.

Now, you can imagine that over time there's a probability that a particular nucleotide mutates, say, twice. So, in the first instance it may change from A to a G, , but then it changes to a C. But when you look at the modern-day sequences, you don't know that this actually happened. And so there's ways to statistically estimate what the likelihood is that a sequence actually contains such multiple events. OK, and this, we called, a corrective evolutionary distance then.

And what you will note is that the corrected evolutionary distance is invariably larger than the actual observed one. Now, what can we can do with those distances? We can constrain them into a best fit tree of relationships. So, we can draw what we call is a best fit tree. That's shown here. We have our four organisms, but when you look at those branches of the tree what you'll see is that they add up roughly to the correct evolutionary distance here. So, between A and B we have 0.

23 and 0.08, which roughly gives you 0.3 here, OK, whereas between A and C the tree is constrain such that we have 0.31, and here 0.15, and so overall you roughly get the distance here that we have calculated. And so what this means is that you ordered the organisms by their calculated evolutionary distance. And so you have now obtained, actually, a very intuitive picture of the relationship of organisms to each other where A and B are obviously the most closely related ones, and A and D are the most distantly related. Is everybody with it? Any questions? OK, now, this best fit tree is what we call a phylogeny.

Now, excuse me, these techniques really revolutionized the study of evolutionary relationships, and one of the things that it allowed us to do is to construct universal phylogenetic trees or what we can also call the tree of life. And I will show you this on the next slide, and that I want to make a few general statements about this. So first of all, when you analyze all known organisms, and obviously that would be a big task, but representative of all known organisms, what you'll find is that, indeed, we have three major lineages: the bacteria, the archaea, and the eukarya.

OK, so we have what we call three domains of life: the archaea, bacteria, and the eukarya. So, this really is the evidence that life really split very, very early on into those three lineages that I showed you before. Interestingly, two of those major domains here are prokaryotic, OK? So, two of the domains are prokaryotes.

Moreover, if you actually look at the types of organisms that are on here, you'll notice that even on the eukaryotic side of the tree, most of the organisms here are actually microbial. So, the single celled organisms: and that means that most of the life on the planet is microbial. The vast diversity of organisms on the planet are microorganisms. So, we can say that most life is microbial. And when you, then, look at analysis of mitochondria, and chloroplasts which all have their own genetic machinery, and therefore also their own ribosomes you'll see that the mitochondrion, OK, and the chloroplasts both tree within the bacteria.

So, we really have an amazing confirmation of this endosymbiont theory which actually developed in the absence of gene sequences by some Russian scientists in the early 20th century. So, we have that mitochondria and chloroplasts tree within bacteria, and this really supports the endosymbiont theory.

So really, you could say eukaryotes are really just walking, and swimming, and flying incubators for bacteria, right? So, just hosts for microorganisms. OK, so basically you can, what you should take home from this is the three domains of life. Two are prokaryotic, and even more so most of the diversity that we find is actually microbial, and then finally the endosymbiont theory is actually confirmed by those phylogenies. Now, what I want to cover in the remaining time, is how we can actually use now those sequences to learn something about organisms in the environment.

That's the topic of molecular ecology. To introduce this, I just want to show you a couple slides that really sort of capture what the big problem is that we're facing here. Now, when we look at the abundance of prokaryotic cells in different types of environments, what we see is that there is an enormous number of different prokaryotes out there. This summarizes, here, different types of environments.

We have the marine environment, freshwater environment, sediment and soils, subsurface sentiments and animal guts. And that this number here gives you the average number of prokaryotic cells either per milliliter or per gram. And it here we have the total number of cells obtained by multiplying the average number with the total volume of the particular environment. So what you can see is that in the marine environment, we have an average half a million cells per milliliter of water, OK? It freshwater, we have about a million cells.

What is that telling you? There's a ton of prokaryotes out there. What you go swimming, you take a little gulp of water: you've probably eaten several million prokaryotes, that it's nothing to worry about because what this also tells us is that very, very few prokaryotes out there are really pathogens because otherwise you'd be sick all the time. Now, in sediments and soils, in as little as a gram you have five times 10^9 prokaryotic cells almost.

5 billion prokaryotic cells are out there, and even in very, very deep sediments that reach down to 3,000 m, you have a substantial number of prokaryotic cells. Well, and here's your guts, 10^5 times 10^6 gives you 10^11 per gram. So again, you're

just a walking incubator for a very complex microbial community. Here's the global abundance. You see that steeps of surface sediments and the marine environment, probably in terms of numbers at least, the most important microbial environments. Now, faced with this enormous abundance of prokaryotes out there, very important question is how many of them are out there? Or, how diverse our prokaryotes in the environment? That's important if you want to figure out their function and the environment, and want to understand also their evolution.

And what I want to show you here is that we've gone through an amazing development in our understanding of prokaryotic diversity in the environment over the last 10 to 15 years or so. Who knows about E.

O. Wilson here? One person? So, he wrote a very famous book on biodiversity, which was published in 1988, where he tried to summarize, really, how diverse the known organisms are on the planet it also try to extrapolate to the total diversity. And what you see is that he came up with about 1.4 million different species here, mostly dominated by insects. That's the big section here on this pie chart. The plants: very important. And if you look, the prokaryotes feature with about 3,500 different species. So, in 1988 we thought there were very few prokaryotic species out there. If you look about 10 years into the future and take the assessment here, and this just exemplifies how the thinking has changed, you see that we think now that there is about 11 million different species out there, and that the vast majority of them are prokaryotic, OK, 10 million.

So, this big part of the pie chart is really the prokaryotic diversity. Now, what really has changed is that we've actually started to use molecular techniques to determine the diversity of prokaryotes in the environment. So molecular ecology is really the use of molecular gene sequences obtained directly from the environment -- -- to learn about the diversity prokaryotic -- -- diversity out there.

Now, this slide just quickly summarizes this. Basically, the idea is that you go out into the environment and collect either water or soil samples that, as I just showed you, invariably contain a lot of different prokaryotic cells. You then lyse the cells and purify their DNA. And so that you end up with a mixture of DNA that represents the organisms out there, and then you can use universal PCR primers to actually amplify ribosomal RNA genes from all the organisms that are present in your samples. Now, why can you use universal PCR primers? Well, they target the regions number A that I showed you before.

Those regions in the genes are invariant amongst all organisms. You guys all remember how the PCR works, right? We cover this. OK? Yes? No? Who doesn't? You don't? All right, come to the board. Just kidding. OK, you should look it up. I don't have time to cover this, unfortunately, but basically it's a technique that allows you to amplify specific types of genes millions to billion fold. And once you have done this, what you can do is that you can purify the genes on gels, and then separate them by cloning them into individual plasmids. And those plasmids have been inserted into E.

coli cells, and the E. coli cells are then individually grown up so that each culture contains only a single plasmid, and you can then sequence these ribosomal DNAs or ribosomal RNA genes from those clones. And so, you have obtained a library of the ribosomal RNA genes from the environment. So, we use environmental ribosomal

RNA gene libraries from which we then can actually compare how many different types of genes are out there. So let me show you an example of this.

What we have done recently, we've gone out in one of the first really comprehensive samplings of coastal bacteria plankton, which means the bacteria that are present free living in ocean water. And so, we've done this, we've collected all those clones, and then basically we constructed those phylogenetic trees that I showed you before that really allow us see how many different types are out there, and how closely related they are to one another. And what we found is that in this environment that you think might be very simple because it just the water column right? No, not much structure in there.

We found over 1500 bacterial 16S ribosomal RNA sequences to occur, so an enormous diversity of prokaryotes of bacteria in that particular environment. And the important point is that when you actually look at a collection of such studies that I just showed you, what you find is that the vast majority of microorganisms in the environment have never been cultured. So traditionally what we do of course to learn about microorganisms when you grow E. coli, or so, you throw them onto culture plates. You make lots of different cells, and that allows you to study some of their properties. But when you look, for example, at results from the ocean, this summarizes now coastal and open ocean environments, again, the bacteria plankton is those free-floating bacterial cells in the water. And you compare this to what we've actually been able to culture from those environments.

What you see is that you have some dominant groups here. They have all funny names, most of them, because they're just clones and clone libraries. But these are the dominant groups that show up in clone libraries. Here's their relative representation in different clone libraries from a variety of environments. And so here you have one very important one, the SAR11 group, or this one, the SAR86, that always show up in clone libraries. But we've never see them in culture, so the important point to realize here is that what is actually happening is that whenever we go out, we find a great diversity of bacteria out there, but we have no idea what they actually do. And this is one of the big questions that we need to answer to understand, really, how the planet actually works.

What are those uncultured microorganisms out in the environment really doing, and what is their importance? And we'll talk about this next time. We're going to talk about environmental genomics because essentially what we can do now, is we have techniques available that allow us to isolate and least large fragments of the genomes, sequence those, and look at what kinds of genes they have present. And that allows us, then, to infer some of their function in the biogeochemical cycles in the environment.

OK, so with this I'm going to close today unless you have any more questions.