MOLECULAR FRONTIERS JOURNAL

Origin of the Eukaryotic Cell*

Nick Lane[1]

All complex life on Earth is composed of 'eukaryotic' cells. Eukaryotes arose just once in 4 billion years, via an endosymbiosis — bacteria entered a simple host cell, evolving into mitochondria, the 'powerhouses' of complex cells. Mitochondria lost most of their genes, retaining only those needed for respiration, giving eukaryotes 'multi-bacterial' power without the costs of maintaining thousands of complete bacterial genomes. These energy savings supported a substantial expansion in nuclear genome size, and far more protein synthesis from each gene.

Keywords: Eukaryotes; Mitochondria; Bacteria; Energy per Gene; Complexity.

Thank you for the very kind introduction and thank you for the invitation, this has been an absolutely wonderful meeting, I have learned a great deal. I should say that I am not an engineer and so I feel as if I am here under slightly false pretenses. What I will talk about though is natural selection, because natural selection is a great engineer. But we heard yesterday from Frances that there are areas of biology where enzymes don't seem to have ventured. It's not that they can't do it, they can do it perfectly well. It's just that there isn't an 'incentive' for them to do it. And we see the same thing with cells and life generally. There are whole areas of possibility that life has just not explored and it's the reasons for that I find fascinating. I'm going to talk today about the origin of the eukaryotic cell. Eukaryotic cells are the large complex cells with a nucleus; and eukaryote just means true nucleus, where we have most of our DNA. Essentially everything you can see, which is to say plants and animals and fungi, they're all composed of these eukaryotic cells.

So this is a three domains Tree of Life that I suspect a lot of you here will be familiar with. It goes back to Carl Woese in 1990[1]. I know a lot of people are still taught today about the five kingdoms. Actually let's just have a quick show of hands from the high school kids here. How many people have heard about the three domains and how many people..? Yeah, okay. And how many people are familiar with the five kingdoms or the six kingdoms of life? Okay, that's great. So I can tell you for nothing that education in Sweden is better than education in the UK, because I've asked this question of UK audiences and very few people have really got beyond the five or six kingdoms. So the three domains are the bacteria, the archaea and the eukaryotes. The bacteria and archaea look more or less the same, which is a key point. We've known about some archaea for hundreds of years. The methanogens for example, we've been familiar with them for 400 years. But they look the same as bacteria in their appearance. Carl Woese first started sequencing RNA and came up with this genetic Tree of Life, and the branch lengths here then give an indication of the amount of variation within these groups. This was all a bit shocking because the animals, the fungi, the plants were compressed into this small corner of the Tree of Life. This was a Copernican revolution in biology because again it pushes us into a small inconsequential corner of the universe, even of life, and it's difficult to accept but it's true. It's also difficult, at least I find it's difficult sometimes, to see things, and this is one of the reasons I enjoy writing books because that's how I come to understand the world. I try and explain, why is the Tree of Life like this? And you don't ask, I didn't ask myself these questions; but there are a couple of very strange things about this Tree of Life. Why

[1]Professor of Evolutionary Biochemistry, University College London. Published online 31 October 2017; doi:10.1142/S2529732517400120

* This article was transcribed from a presentation delivered by Dr. Lane at the Molecular Frontiers Symposium on Tailored Biology at the Royal Swedish Academy of Sciences, May 2017. It is published here in *MFJ* with permission from Dr. Lane.

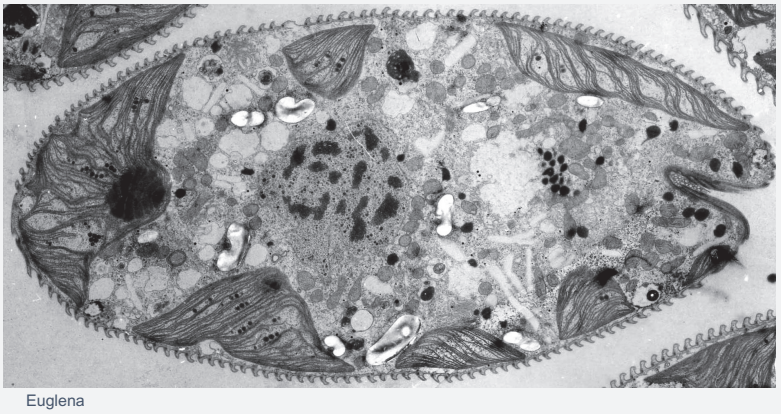

Figure 1. Difference in scale between a single-celled eukaryote (the alga Euglena) and a relatively complex bacterium with an intracellular compartment (planctomycetes). Roughly to scale. Photomicrograph of Euglena courtesy of M. Farmer.

is there so much genetic variation within these two groups? They're very different to each other, they're different in their biochemistry and all of these metabolic pathways that we've just been hearing about, they are very different, shockingly different. But in their morphology they're very similar. They've had four billion years of evolution to come up with anything in their biochemistry. It seems like they are far more sophisticated than the complex eukaryotic cells over here. They can do anything except become large and big and complex. Why didn't they do that? What was happening down here that wasn't happening in those groups? And it becomes even more strange because it's not just at the level of large plants and animals, it's at the level of single cells. So this is *Euglena*, basically it's the kind of scum that you find on any pond (Figure 1). Here is one of the more complex bacteria that you'll find. This is planctomycetes, you might just be able to see that it's got a little…looks a little bit like a nucleus. It's a kind of compartment where the DNA is, it's not very much like a nucleus, but this is one of the reasons we consider it to be quite complex; some people suggest that perhaps this is a first step towards making a nucleus. But the reason you can't see it very well is that this is roughly to scale. This is just enormously larger — on average eukaryotic cells are about 15,000 times bigger than bacteria in their volume. And you don't need to know what these are, these are actually chloroplasts in the case of *Euglena*. Here's the nucleus, you can just make out the mitochondria, but you don't need to know what this stuff is inside a cell to appreciate that we've gone up orders of magnitude in complexity from what the bacteria and the archaea have done, and it's a puzzle.

Also at the level of eukaryotic cells, again just to make you feel even more inconsequential, this is a paramecium and this is a pancreatic acinar cell. I'm curious to know how many genes do you think a paramecium has? When I say genes, I mean genes coding for proteins. Any suggestions for how many genes paramecium has? I've had answers between 50 and a few thousand, how many do you think? Okay, well, I'll tell you the answer is 40,000. That's twice as many as we have. These are complex cells, but single-celled organisms. The reason they have so many genes is they do an awful lot of stuff inside that single cell. So, the level of complexity between different types of eukaryotic cell is very equivalent. We are really not very much more sophisticated than paramecia it seems, but we are an awful lot more sophisticated than a bacterium in terms of the morphological complexity.

So, what's going on? There's really no agreement so I'm confident in standing up here and talking to you, that whatever I say may be right or may be wrong but you'll never really know. But I think we can try to get at the problem even if we don't know the answer. These are some of the really great thinkers of biology over the 20th century. Jacques Monod we heard about yesterday, he was one of the pioneers of molecular biology. He wrote a wonderful book called Chance and Necessity in 1970. It's got quite a bleak existential view. It's an extremely exciting read but he sees the origins of life as being really very, very difficult and unlikely to happen again. One of the reasons that perhaps we don't see life elsewhere is that the origin of life is so difficult. That's what he thought. I don't think the other people in this room would necessarily think that way. But again, once evolution has got going what happens? Stephen Jay Gould wrote a great book called Wonderful Life where he imagined winding back the clock to the time of the Cambrian explosion when the first animals appeared in the fossil record, and then let it play forward again; would we end up with humans? Would we end up with even vertebrates? Or would we end up with giant octopuses on the hills or something? It's very hard to know what you might imagine. The question is what kind of engineering principles guide the evolution of life? So, on this side, these two basically think that there's a great deal of contingency, that the environment affects what happens, that the asteroid that wipes out the dinosaurs gives the mammals a chance that would never have happened if an asteroid hadn't hit

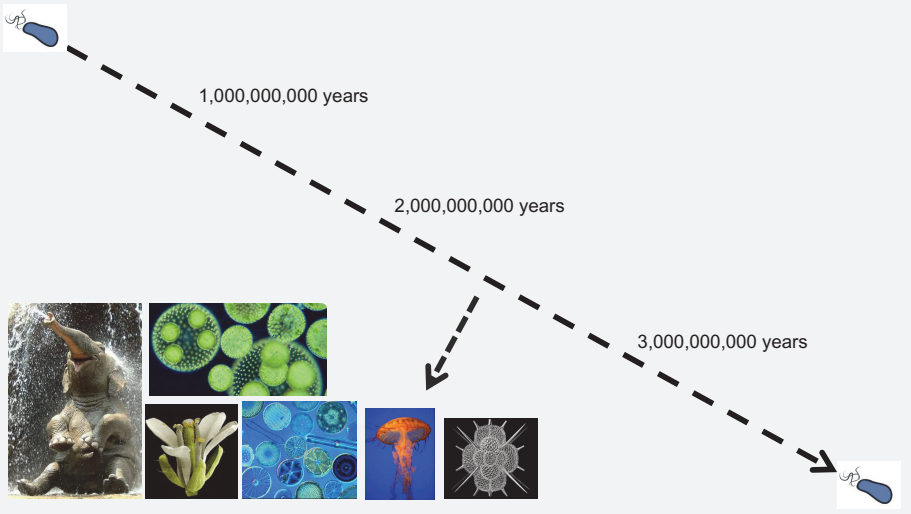

Figure 2. An unusual representation of the tree of life: in terms of their morphological complexity bacteria and archaea have barely changed over 4 billion years; morphologically complex eukaryotes, in contrast, arose from a singular endosymbiosis between two prokaryotes around 1.5 to 2 billion years ago.

in that particular place — in fact it seems perhaps vaporizing sulfates into the atmosphere by hitting specifically there in the Yucatan Peninsula. Christian de Duve, another Nobel laureate, and Simon Conway Morris believed far more in the engineering principles that underpin selection, that you will get the same things emerging time and time again because that's the best way to do it. If you want to fly you'd better have something like wings, you'd better be aerodynamic. You're going to find wings arising in bats, in birds, in insects with a rather similar structure, with rather similar aerodynamic properties for the same reasons. And so, they would argue that we would end up with something rather like the world that we have at the moment, if you were to wind back the clock to the origin of life and let it play forward again. The simple answer is we don't know who is right.

This, I think is the key problem. And I've already alluded to this. If we go back nearly 4 billion years, the kind of timeline that Gerald Joyce was talking about in the in the last talk, 3.6 to 3.8 billion years ago we see fossils in rocks that look a lot like bacteria. We don't know for sure if they are. Natural processes can give rise to shapes in rocks that look an awful lot like bacteria, but to the best of our knowledge, we see bacteria going back 3.8 billion years. A recent paper from colleagues of mine at UCL said 4 billion years[2]. I'm not sure if they're right. But, look what happened afterwards, they flat-lined for practically three and a half billion years (Figure 2). What was going on in, or what was not going on in the bacteria? Why is it that only once in this entire period do we see the origin of complex life? Now we know it happened once because all these eukaryotic cells are related to each other. We share a tremendous number of traits in common. We all have a nucleus, we all have straight chromosomes, we all have genes in pieces with non-coding introns and then bits that code for proteins, we all have the same structures inside cell, we all have mitochondria and I'll say more about that. So, we know that we all share a common ancestor. And by definition that arose once; but if we look in the fossil record, we do not see things that were not eukaryotic, other origins, alternative origins of complex life. We don't see it there. We look around the world, we trawl through all kinds of muds in strange environments and we look to see alternative forms of life and we don't really find it. We find new archaea, we find new bacteria, we find amazing things, but we're not really finding different structures to cells[3].

This is one of the great evolutionary biologists and he was at UCL for a period — John Maynard Smith. He used to look for what he called the scandals of evolution. The things that really ought not to be happening like that, they should have done something different. Why did it go this way rather than that way? And this is an evolutionary scandal by his terms — all complex life is composed of eukaryotic cells. They only arose once and we all share not just the physical structure of the cells but we are all sexual. Plants are sexual and yeast are sexual too. It is right across the entire tree of eukaryotes, not just sexual, but the gametes fuse together and they go through a two-step meiosis using the same proteins, we can find the same genes right across the whole tree. It's the same. Why don't bacteria evolve any of these complex traits? They do some homologous recombination and lateral gene transfer, but they don't do two-step meiosis and they don't recombine across the entire genome[4]. It's a very different process that they do. So, the scandal is, if all of these traits evolve step by step by natural selection and each step offers some small advantage (and there's no reason to disbelieve any of that) then why is it that none of these traits arose in bacteria? It ought to be like the eye. Eyes arose essentially independently on at least 60 or 70 different occasions in different environments. A lot of these are animal eyes and so they actually do go back to a common ancestor that was a light sensitive spot on

some kind of a worm, and there are some regulatory genes that they have in common, a PAX6 gene for example, but independently those regulatory genes recruited all the rest of the genes required to make an eye. And so, the octopus eye, which is here, and the human eye, they are very fine examples of convergent evolution. They're structured in essentially the same way but they evolved independently. This is *Euglena* again, here's the eye spot in *Euglena* that uses essentially the same rhodopsins that we use in our own eyes. And this is even more strange — this is a single-celled protist and its got a retina here, and its got a lens, its got a cornea. The retina is made of chloroplasts. The cornea is made of mitochondria. Its just recruited those different parts. Its got the same structure as the kind of eye that we're familiar with, but it's an utterly independent origin, this is convergent evolution. Selection would predict that we should see multiple origins of rather similar functioning things that are different to each other in different environments, different ecosystems. So why don't we see multiple origins of a nucleus, if it's a good thing to have, or sex if it's good to have sex, or phagocytosis, the ability to go around and engulf other cells (essentially eating), but we never see that in bacteria.

Well, here is a way of getting at the problem. This is a more recent tree[5]. It's actually a few years old now but I like this one. This is just the eukaryotes. And again forget about plants or algae as the main groups. There are five or six super groups of eukaryotes. There are the excavates, the chromalveolates, the unikonts, we are all unikonts. I rather like the term although it's now becoming slightly old outmoded but we are all unikonts. Here we are, the metazoans and the fungi. The reason I like this particular tree is right at the center: this is the common ancestor of all eukaryotes, and it's rather symbolically a black hole. So there are two things to take away from this. There's far more variation within these groups than there is between the ancestors of the groups. This is a what's called a Big Bang radiation. It happens apparently rather abruptly. And that common ancestor had everything, it was a recognizable eukaryotic cell. We can trace the genomes of a large number of cells in these different groups and we can see that they share an awful lot including all these traits that I've been talking about. So we know that the eukaryotic common ancestor had all of those things, but we don't know how they arose. What we know is that bacteria don't have them. This is what I like to think of as the black hole at the heart of biology. We do not know how or why complex cells arose from bacteria. So I will put some ideas forward…these are not by any means the only ideas and we don't have facts to prove anything yet. But this looks like some kind of a bottleneck, it looks as if perhaps the conditions changed and the reason that eukaryotes suddenly took over the world is perhaps there's been a snowball earth, we know there was a snowball earth about 2.3 billion years ago and another one around 700 million years ago. This is when the entire planet froze over, we think, the geologists tell us, right down to sea level on the equator. Catastrophic global changes are undoubtedly bottlenecks that could affect tremendously the whole trajectory of life. This is another one, the Great Oxidation Event when we first start to see oxygen in the atmosphere, again from about 2.2–2.3 billion years ago, probably linked with that earlier snowball earth. These are global catastrophes and it is very easy to imagine that after this catastrophe, just as the mammals expanded after the dinosaurs, the eukaryotic cells expanded.

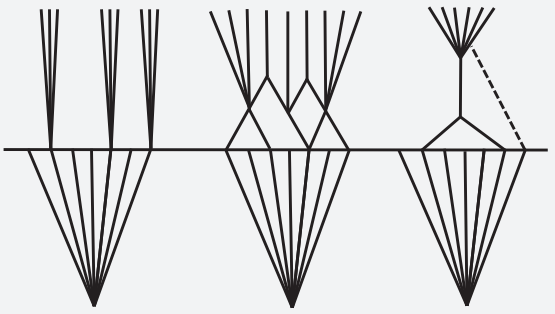

Figure 3. Some predicted schematic trees. Left panel: If complex cells arose in response to a change in environmental conditions, such as oxygenation, one would predict polyphyletic origins of complex cells, in which photosynthetic bacteria gave rise to algae, osmotrophic bacteria to fungi etc. Middle panel: The prediction of polyphyletic origins would also be true for serial endosymbiosis, in which different symbioses occurred in disparate environments. Right panel: the puzzling reality — a singular endosymbiosis gives rise to a monophyletic origin of eukaryotes, with no surviving evolutionary intermediates. The dotted line depicts the later endosymbiosis that gave rise to chloroplasts, but this did not affect the origin or early evolution of eukaryotes.

But that makes some predictions and people have been a little sloppy in the way that we've thought about these predictions (Figure 3). Well you might imagine if it was oxygen, for example, suddenly allowing the freedom to become bigger and more complex because now there's oxygen that animals and plants can respire and so on. You would expect multiple origins nonetheless, you would expect that the cyanobacteria, the photosynthetic bacteria, would give rise to photosynthetic plants and algae. You would expect that osmotrophic bacteria would give rise to fungi, putting enzymes out into the surrounding area, breaking down food and then taking up the monomers. You would expect separate origins from the cells that are best pre-adapted to the new conditions, but that's not what we see. The Serial Endosymbiosis Theory from Lynn Margulis again, she anticipated that there would be multiple different types of endosymbiosis[6], different types of bacteria interacting with each other, and so multiple different origins of complex life, but that's not what we see[7]. What we seem to see is something more like this, and I shall

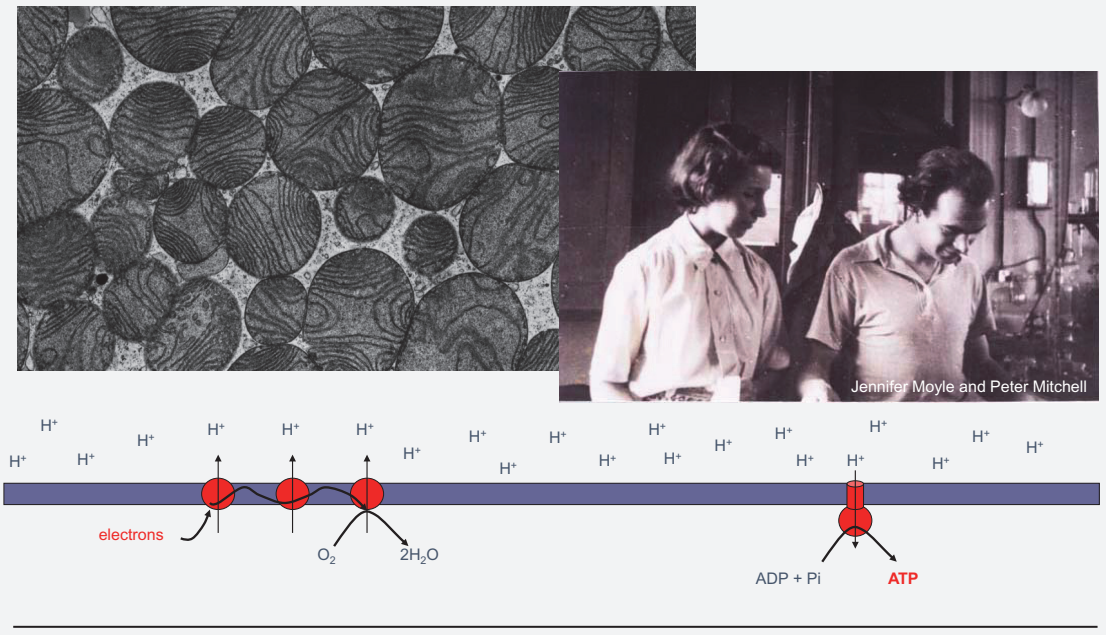

Figure 4. Mitochondria showing the cristae membranes where respiration takes place. The mechanism of respiration, known as chemiosmotic coupling, was elucidated by Peter Mitchell in collaboration with Jennifer Moyle and others. The cartoon shows the mechanism of chemiosmotic coupling, in which the flow of electrons from 'food' to oxygen within the membrane powers the extrusion of protons across the membrane; the flow of protons through the ATP synthase, a kind of turbine, powers the synthesis of ATP, the universal energy currency of life.

say more about this. It does not look like an environmental bottleneck but a structural bottleneck, something to do with the structure of cells, where one cell got inside another one (Figure 3). This is the black hole area where everything evolved, and this is the moment where everything takes off into the modern groups that we see. This is the acquisition of the chloroplasts here — the cyanobacteria that gave rise to the algae. They didn't change the fundamental direction of evolution. But here is what I shall talk about for the remainder of the talk.

The one thing which has really changed over the last 20 years, and again this is grounded in medical research because most of these cells, once called archezoa, are in fact parasites of one sort or another. It was thought that none of these cells have mitochondria, and they all looked rather morphologically primitive. The assumption was that the archezoa were early branching eukaryotes that would give us an indication of how complex life arose. Well it turns out after a lot of studying that they all do have mitochondria, just not as we know them. They've become what are called relict organelles, although they were lost altogether in the case of *Monocercomonoides*[8], but they all had them once and they became specialized for different tasks. So we now know that the common ancestor of eukaryotes already had mitochondria. And when we start looking at the genomes then we can see that potentially the origin of eukaryotes and the acquisition of mitochondria were one and the same thing. The mitochondria, in case you don't know, they are the powerhouses of eukaryotic cells. These are our own mitochondria here. And what's going on here was discovered back in 1961 by Peter Mitchell, who called it chemiosmotic coupling[9]. This is Peter Mitchell in 1947 with Jennifer Moyle, his long term collaborator, through all their lives really (Figure 4). Mitchell won the Nobel Prize in 1978 for his visionary ideas. Jennifer Moyle had been the experimentalist who tested most of these ideas and showed that they were essentially true, that this really is how cells work. It's interesting to me, perhaps in this arena especially, to wonder about how one balances between the ideas and the experiments. Mitchell received the Nobel Prize because he had developed the ideas himself, but he didn't really do the experiments. If Jennifer Moyle had not done those experiments nothing would have proved Mitchell to be correct, and so perhaps they deserved it together, I don't know. But this is what they showed, this is what's happening in you right now. Electrons are being stripped from food and they're being passed down the respiratory chain (Figure 4). This is the membrane here, the cristae membranes in here. These are giant protein complexes of the respiratory chain and I'm just symbolizing them as small balloons. What we have is a current of electrons from food to oxygen, and that current of electrons is powering the extrusion of protons across the membrane. So we end up with a kind of reservoir

of protons on this side of the membrane and relatively few on that side. Protons have a positive charge which means you now have an electrical potential across this membrane. But there's also a concentration difference. This is what Peter Mitchell called the proton-motive force. And this, the ATP synthase, is a wonderful machine, it's a rotating motor, essentially equivalent to a turbine in a hydroelectric dam. This is equivalent to water flowing through the turbine: the protons flowing through the ATP synthase are turning the motor, which is powering the synthesis of ATP, and that's powering everything else in the cell. So this is what's happening in respiration. Now the mitochondria were bacteria once, we're pretty certain about that, in fact there's not really any serious opposition to that idea anymore. That goes back to Lynn Margulis, in fact it goes back much before Lynn Margulis, but she was the person who nailed the idea in 1967, 50 years ago this year[6]; there's a celebratory issue of The Journal of Theoretical Biology where she published that original paper coming out shortly. Incidentally, she was married to Carl Sagan — they must have had some pretty amazing breakfast time conversations, I should imagine, in that household. They unfortunately divorced before 1967 when she first published that paper. But the key point is that mitochondria were bacteria once, and that's firmly established.

This is an alternative Tree of Life going back to 1992, about the same time as Carl Woese and this is a different way of seeing it[10]. This is Jim Lake. So this is the Carl Woese' Tree of Life, the three domains Tree of Life. And Jim Lake said no, no, it's not, it's a ring; apparently he sent a paper to Nature entitled One Ring to Rule Them All. And they rejected the title unfortunately but published the paper, and this was the essence of it. He was looking at where these genes came from. And a large number come from bacteria, a large number come from archaea, and it's this genomic fusion that is giving rise to the eukaryotes. This was a radical idea that nobody really believed for quite a long time but over the last maybe six or seven years it's become clear from phylogenetics that something like this is indeed the case. So this is the classic three domains tree with the eukaryotes at the top. These are different groups of archaea here and the bacteria down at the bottom. That's the classic tree. What we see now is that the eukaryotes branch inside the archaea[11]. This is from concatenated sequences of 40 to 60 genes, the more genes you have, the stronger the signal, the trouble is the more genes you have the more likely to be passed around by lateral gene transfer they are, and that produces noise which confounds the signal. So it's a difficult balance but this has been repeatedly found in a lot of studies now. All this means the host cell was an archaeon. We don't know what kind of archaeon but we're getting closer. This is a paper from earlier this year and some earlier work from a couple of years ago. This is the Asgard superphylum. The Lokiarchaeota were discovered at Loki's Castle a couple of years ago[12]. We don't know what they look like. This is just metagenomic screening of the muds around there. And these are the most similar genomes to eukaryotes — here are the eukaryotes branching. There's now several groups so we have the Lokiarchaeota and then the Odinarchaeota, the Thorarchaeota, the Heimdahlarchaeota, we've got all these Nordic gods. We don't know what any of these cells look like, we only have the genome sequences and they're all relatively small, they're all 4,000–5,000 genes or so. So they're kind of standard size for archaeal genomes. And they have some interesting properties. They seem to have a pretty dynamic cytoskeleton compared with other archaea. They seem to be capable of some membrane remodeling, but we haven't seen it. We don't really know if are they slightly phagocytic, can they begin to engulf other cells, or is it just how they divide in half where you also need to change the membrane structures. So, we don't know yet what kind of archaea, but we are fairly sure that we have something here, this, something within the Asgard phylum, the Lokiarchaeota acquires a bacterium.

And this is where it becomes really quite difficult and I'd just like to show you. This is a paper from last year from Nature which gives an indication of how difficult it is to get at some of these problems[13]. This is looking at the stem length of genes that come from bacteria. There's a lot of genes that come from bacteria, some of them branch with the alphaproteobacteria. These are definitely the mitochondrial genes. Here are the genes that are definitely with the mitochondria. But these genes branch with the bacteria but not clearly with any particular group. And the stem length is longer which means to say there are more differences. This structure is the nucleus, these are bacterial genes that you find in the nucleus, and they are more distantly related. These are the endomembrane systems, these are the mitochondria. So what they said is that the number of differences represent time roughly, basically a molecular clock. There's an assumption that the number of differences accumulating over time gives an indication of the time that has passed. If that's the case then the nucleolus, the nucleus, the endomembrane systems all arose before the acquisition of mitochondria. And if this is correct then everything I'm going to tell you in the rest of the talk is incorrect. That's the kind of thing that troubles you when you're trying to sleep at night. Are they right? Have I been wasting my time for the last 10 years? Well, they may be correct, I don't know. They may also be wrong. The question is, do I have an explanation as to why they would see that. I assume their data is correct (though I know it has been challenged). So, do I have an answer? Well, yes, I do; I don't know if my answer is correct but it helps me to sleep. Let me explain.

This is an old tree going back now to 1998[14]. Here are the bacteria, the archaea and these groups, these are the

cells that I showed you before, the Archezoa. These are cells that don't have mitochondria. We used to think that they were early branching. Why? Well, they're simple in their morphology but also they have these long, long branch lengths. And that means that the place they branch on the tree is an artifact called long branch attraction. They're shown to branch here and here was the acquisition of mitochondria, actually right up there somewhere. So the length of these branches does not give an indication of the amount of time, it gives you an indication of the amount of evolution that has happened, the number of changes. And we can't constrain the time necessarily with that. So, this tree is an artifact and everybody's agreed about that now. And I think that is probably the best way to think about this other study.

This is a painting by Odra Noel, a beautiful painting of a eukaryotic cell. Here is the nucleus, here are the mitochondria, here are these endomembrane systems that I was talking about. Why would you have more evolution in the endomembranes or the nucleus? Well, the genes in the mitochondria are doing what they always did. They're doing respiration in a mitochondrial setting. They are under strong purifying selection for the same job in the same setting that they always did, it never changed. And purifying selection means that you have fewer changes in sequence because changes get eliminated if they're not helpful to you. But the endomembranes, well, they don't exist in bacteria really. And the nucleus, it doesn't exist in bacteria. So there must have been, theoretically there must have been a period of strong adaptive selection and adaptive selection by its nature is forcing changes on you. You're changing to a new purpose, a new function and so you're going to have lots of changes. And that's going to increase the branch length.

If I'm right, what does it say? Well, this is the origin of the eukaryotic cell. These are bacteria living inside a bacterial cell (Figure 5). This is the only example that we know of, of bacteria inside a free-living bacterial cell[15]. There's plenty of bacteria inside eukaryotic cells which are large and complex and often engulf cells like bacteria for a living; but bacteria don't do that. This is a cyanobacterium, these are thylakoid membranes, it has a cell wall. It did not engulf those cells by phagocytosis, but we don't know how they got there. But what good is it if those cells inside went on to become the mitochondria? Why was the acquisition of mitochondria any use? It seems reasonable that we should look for the answer in terms of energy in one way or another. Well, they are the power packs, they produce the ATP so that's where we should look for an answer. But, if you look, these are values taken from the literature[16]. This is the metabolic rate of bacteria compared with single-celled protists. This is a log scale in each case so the bacteria respire about three times faster than single celled eukaryotes. So, it's not the case that they help us to respire faster, it's not as simple as that; but that's

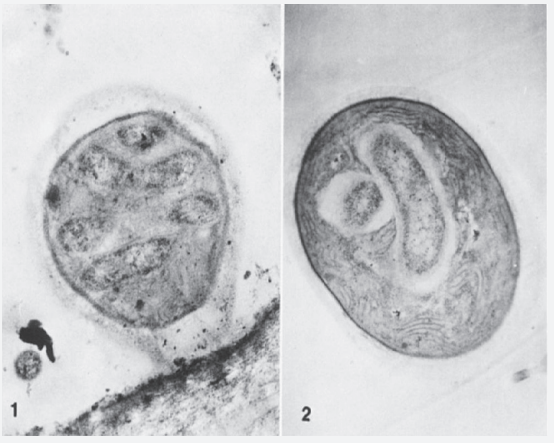

Figure 5. Bacterial cells in the cytoplasm of cyanobacteria, the only known example of intracellular bacteria within free-living bacteria. The endosymbiosis at the origin of eukaryotes might have looked a little similar to this. Courtesy of D. Wujek.

per gram. If we look per cell, it changes around. So, this is a log scale again. Now here we can see that a single eukaryotic cell consumes about 5000 times more oxygen per minute than a single bacterial cell[16]. Why? Because they're 15,000 times larger — of course they do, they're juggernauts. This is a silly comparison in one sense, but it's beginning to get at the problem. What are they spending all this energy all on? Well, this is energy per mega base of DNA. And you can see it's roughly similar again in this case. So, eukaryotic cells are becoming a lot larger, they have a lot more energy to spend and they're spending it on maintaining a much larger genome; but they're not really spending any more per mega-base of DNA than in a bacterium does. What are they actually spending it on? Well, this is some old work. This is not the kind of thing that many people think about anymore, but Frank Harold did some lovely work in the 1970s on the ATP budget of bacteria[17]. The answer is to a large degree protein synthesis. That's not true of us, that's not true of multicellular organisms generally. But it seems to be true of many protists and bacteria that 75 to 80 percent of the energy budget of a cell goes on protein synthesis. Here is DNA synthesis, it's a trivial cost in comparison. There can be lots of futile cycling going on in an environment but this is in growing cells. And when they're growing the biggest costs are to do with protein synthesis. So if we then look at the energy availability per protein coding gene and we equalize for the number of genes then we find again a roughly 5,000 fold difference (Figure 6). That's to say, a single eukaryotic cell has about 5,000 times more energy per gene than a single bacterial cell. That does not mean that they should have 5,000 times more genes, it means that they can do 5,000 times as much gene expression. They can make 5,000 times as many proteins — because they're 15,000

Figure 6. Energy availability per haploid copy of a gene, corrected for genome size (left) and cell volume (right). Data from Lane and Martin (2010), based on measured metabolic rates and mean genome sizes but scaled on the basis of an assumption that bacteria approximate to a sphere and respire across their plasma membrane only. Note the log scale on the Y axis.

times larger of course they have to. It's a statement of the obvious really. So we need to correct for cell volume as well. But now if we correct for cell volume then we get this massive difference opening up — this equates to a 200,000 fold difference in energy per gene when we've corrected for the gene number and cell volume (Figure 6)[16]. Now really this is silly because we've made a silly assumption underpinning it, but it's interesting to get it why it's a silly assumption. So bacteria pump protons in exactly the same way that our mitochondria do. They're pumping them across the plasma membrane. And if you increase the volume of the cell, then of course you have surface area to volume constraints. ATP synthesis depends on the surface area, and protein synthesis depends on the volume. That's where that number of 200,000 came from, and it's a silly number because we know that bacteria can internalize membranes and get around that problem immediately (Figure 7). So why don't they actually do that; or do they do that? Well, to an extent they do. Let's see what the problem is.

Here is another paper that seems to be at odds with my position. This is from Mike Lynch[18]. And what he's plotted here is the number of ATPase enzymes against the surface area. You can see a nice straight line, bacteria down here, eukaryotes up there. So it basically correlates beautifully with surface area — exactly what you would expect. You have more surface area, you have more energy, more ATPase enzymes. And this is the number of ribosomes in a cell against the cell volume. Again, the bigger the cell, the more ribosomes. So you would say that these are just continuous and really there's nothing special about eukaryotes. But what's concealed here in these log scales is that all the bacteria are down here, in every case; we've got one eukaryote there but by and large all the eukaryotes are up here, all the bacteria down there, and there's a couple of orders of magnitude difference between the largest bacteria and the smallest eukaryotes, there's practically no overlap. I would say these are two different continua, and there is something else stopping bacteria from expanding their surface area up to eukaryotic proportions. Look here (Figure 8). These are bacteria: cyanobacteria,

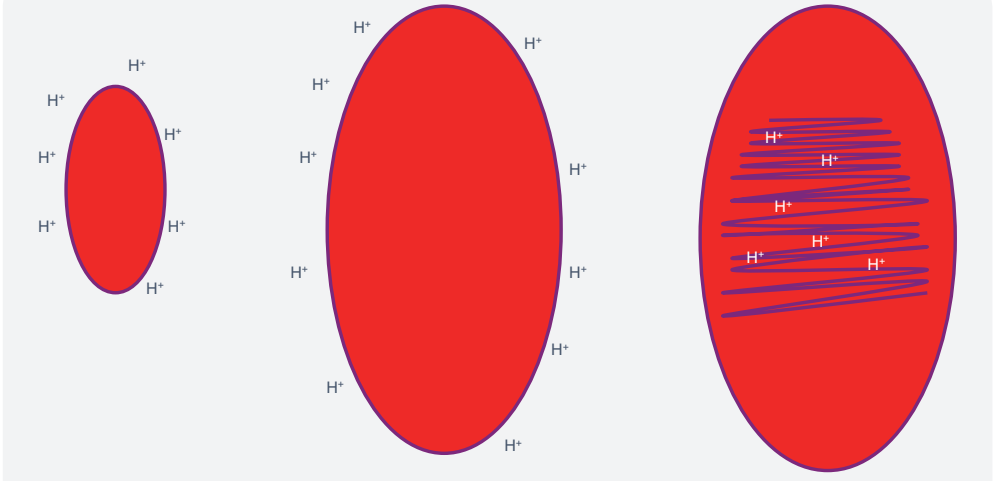

Figure 7. The scaling assumption — bacterial cells respire across their plasma membrane, so ATP synthesis scales with the surface area but the requirements for protein synthesis scale with the cell volume. The size of bacteria should therefore theoretically be constrained by surface-area to volume constraints. However, internal membranes (right panel) which are found in many bacteria, should escape these scaling constraints.

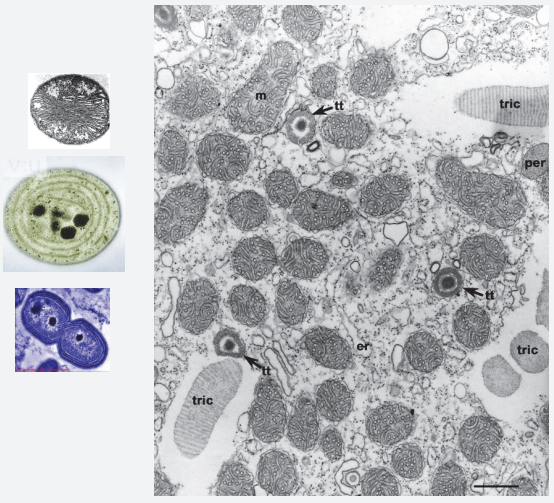

Figure 8. Complex internal membranes in bacteria (left panel; nitrosococcus, cyanobacteria and nitrosomonas) compared with mitochondria in a section of the single-celled eukaryote Paramecium (courtesy of R. Allen). Roughly to scale. Eukaryotes do not only have internal membranes, they also have pared-down bioenergetic genomes, the mitochondrial genome. They therefore display a topological restructuring in which bioenergetic membranes are invariably associated with small bioenergetic genomes that seem to be necessary for their function.

nitrogen fixing bacteria; and these are the internal membranes — they can internalise membranes perfectly easily. And this is paramecium again — these are the mitochondria in paramecium. The difference is scale again. These are roughly to scale. And this is only a small section of a paramecium. It's expanded up over orders of magnitude. The difference here is that these mitochondria all have genes of their own. They started out as bacteria, they lost most of their genes, but they ended up retaining always a very similar subset of genes. They seem to need those genes to control respiration. And so here we're going to converge on engineering principles in one sense. You need genes to control respiration[19]. If you want to get bigger then you just have more mitochondria. Each one comes with its own regulatory system in these genes. So if they want to expand up to eukaryotic size, well perhaps bacteria can't because they don't have the genes. That's easily tested because there are some giant bacteria around[20]. This is *E. coli* here, here's paramecium. Paramecium is dwarfed by this battleship of a cell. This is *Epulopiscium*, it's a bacterial cell and you can see it with the naked eye (Figure 9). This is even larger — *Thiomargarita*. It's basically a giant vacuole with a thin film of cytoplasm surrounding it; this is a single bacterial cell and this is *Drosophila* the fruit fly. So *Thiomargarita* is almost as big as the head of *Drosophila*, it really is a monstrous cell. If you need genes to control respiration and this cell is respiring across this plasma membrane then there better be a lot of genes there, otherwise these ideas are just wrong. Well, this is known as extreme polyploidy. This is *Epulopiscium*, this is DAPI staining and these are copies of the complete genome (Figure 9).

When I first saw this picture I realized that maybe there's something in what we're talking about. There are 200,000 copies of the complete genome[20]. Each genome is three mega bases of DNA, so 3 million base pairs. And this is Thiomargarita, this is the giant vacuole, this is the thin film of cytoplasm, and here there's about 15–20,000 copies of the complete genome. We can add up all of that genomic weight and ask what's the energy available per gene per haploid copy of each gene. This is *E. coli*, *Thiomargarita*, *Epulopiscium*: the energy per gene is essentially the same. What they are is a kind of a consortium of bacteria that are fused together, and

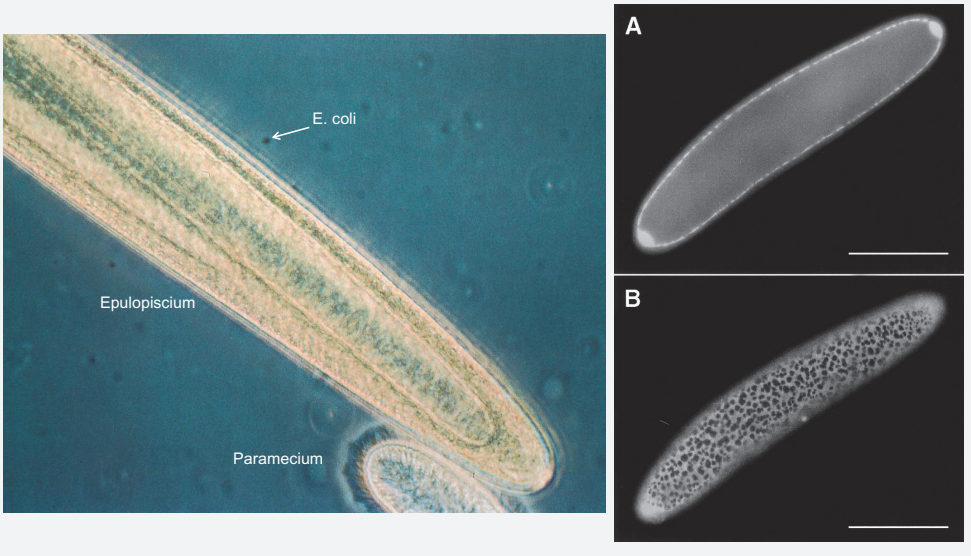

Figure 9. An example of a giant bacterium, Epulopiscium, with extreme polyploidy (shown with DAPI staining, right panel). Epulopiscium can have as many as 200,000 copies of the complete genome, positioned right next to the bioenergetic plasma membrane, as predicted. Extreme polyploidy has a huge energetic cost. Courtesy of E. Angert.

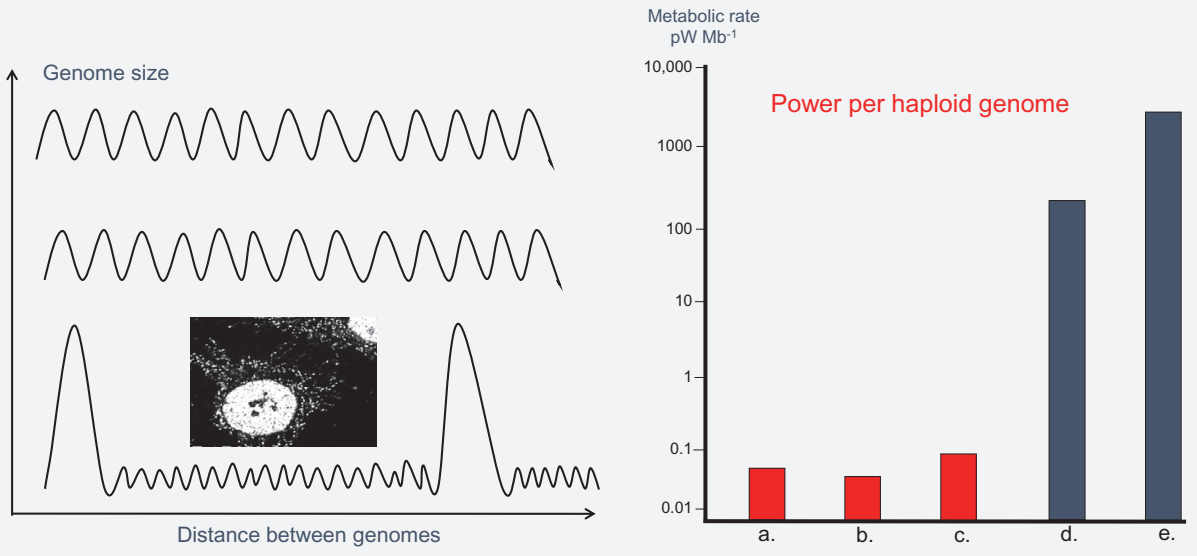

Figure 10. The cost of extreme polyploidy in comparison with mitochondrial genomes. Bacteria have a genomic symmetry (right panel) in which each genome controls a similar volume of cytoplasm in both small bacteria (E. coli, top) or giant bacteria (Epulopiscium, middle). In contrast, eukaryotes have a genomic asymmetry in which tiny mitochondrial genomes (inset, bottom) support, energetically, a massive nuclear genome. Taking into account these differences shows that eukaryotes have orders of magnitude more energy availability per haploid copy of each gene (right panel). Data from Lane and Martin (2010), based on measured metabolic rate and known genome sizes and ploidy. Bacteria shown in red, eukaryotes in blue; note the log scale on the Y axis. (a) E. coli; (b) Thiomargarita; (c) Epulopiscium; (d) Euglena; (e) Amoeba proteus.

each genome is controlling a similar volume of cytoplasm and a similar area of plasma membrane. No doubt there are other advantages to being larger but they have got nothing to do with energy. So why then are eukaryotes different, why is a cell within a cell an advantage? Well, these cells went on to become the mitochondria and the difference is they are not genomes (Figure 10). They are bacteria, they are autonomous self-replicating cells in populations capable of undergoing selection and change over time, and that's exactly what happened[16]. What happens in bacteria generally? Well, imagine the yellow cell here is a cell that lost a gene that it doesn't need. Let's say it's for respiring with lactose or something like that. It doesn't need that gene and if it loses it, it cuts out a bit of DNA. It will grow a little bit faster and there's plenty of evidence showing that gene loss is an important factor in bacterial evolution. Over time most of the cells in that population will lose that bit of DNA that they don't really need because now they can grow slightly faster. But then the conditions change. Suddenly the environment is swamped with lactose again so the cells need that gene again. What do they do? Well, they pick up random bits of DNA from the environment, and one of these happens to contain that gene, and before you know it you're back where you started. And so we have opposing selection processes. One is to lose as many genes as you can afford to lose and the other is to pick them up whenever you need them by lateral gene transfer. But now imagine the same thing is happening inside a cell. Let's say this cell loses the genes required for making a cell wall and you don't need a cell wall if you're living inside another cell, you're in a fairly unchanging homeostatic environment. So the descendants of this cell begin to dominate and eventually all the cells in the population lose that gene, perhaps the genes for a bacterial flagellum and so on as well. You don't need those genes anymore so you lose them and the conditions never change. As long as the host cell survives, the conditions don't change and you do fine.

This trajectory is really common in bacteria. This is typhus, which is a nice example. Typhus obliterated the Napoleonic armies — this is the retreat of Napoleon from Moscow. And this is *Rickettsia*, the cause of Typhus. It is transmitted by the flea and has lost most of its genes — it's now down around about one mega base of DNA, so about a quarter of the size of *E. coli*. That trajectory is really common. This is the range of genome sizes of free living bacteria, which goes up to about 12 or 13 mega bases of DNA. Again, there is a continuum with eukaryotes, but eukaryotes go up to 150,000 mega bases, so orders of magnitude more[21]. And the obligate symbionts and endosymbionts are down here, one mega base or less. We know hundreds of examples of bacterial cells living inside eukaryotes and virtually all of them have undergone this genomic streamlining. Why is that useful? Well, it's useful because the bacteria still make as much ATP as they ever did,

Energy savings	ATP cost	Actin costs	ATP cost
5% of 4 Mb genome = 200 proteins	200	Length of monomer	29 nm
100 endosymbiont genomes	100	Monomers per micron	35
2000 copies of each protein	2000	374 residues per monomer	374
250 amino acids per protein	250	Dimers in actin filament	2
5 ATPs per peptide bond	5	5 ATPs per peptide bond	5
Total per 24 hr lifecycle	50×10^9	Total per micron of actin	131,000
Total per second	580,000	Microns per second	4

Figure 11. Table showing energy savings as a result of 5% gene loss from 100 endosymbionts (left column) and scale of advantage in terms of de novo synthesis of actin filaments, the basis of the dynamic eukaryotic cytoskeleton. The energy savings from not needing to synthesise proteins from 200 genes in 100 endosymbionts equates to 580,000 ATPs per second, assuming a 24-hour lifecycle; in principle this is enough to power the de novo synthesis of 4 micometers of actin per second.

it's just that their overhead costs are being constantly lowered. Imagine that you've got a hundred endosymbionts. This is a silly thought experiment but it gives a gives an indication of the size of the advantage[22]. So imagine you've got a hundred endosymbionts, a hundred bacteria living inside this cell, and each of them has a standard bacterial sized genome with four mega bases of DNA, so about 4,000 genes (Figure 11). And let's assume that they lose 200 of these genes, which they don't need. What are the energy savings of not making those proteins? Well, it is losing 200 genes from 100 endosymbionts, and in bacteria each gene would normally produce about 2,000 copies of the protein that it encodes. On average in bacteria, there's about 250 amino acids in a single protein and the ATP costs are round about five ATPs per peptide bond. So we have a total cost of 50 billion ATP to make those proteins' or the equivalent cost savings if we don't make those proteins. If you translate that into a 24 hour life cycle that would be 580,000 ATPs per second of energy savings. What could you spend it on? Well, imagine a dynamic cytoskeleton, an actin cytoskeleton, which is one of the things that sets eukaryotes apart. What are the costs of that? Well actin is made up of a series of globular proteins which are joined together into a filament. And there's two filaments wrapped around each other. The length of the monomer, the single globular protein subunit, is 29 nanometers, which means there are 35 of them in a micrometer. There's 374 amino acids in each of these monomers and again we assume 5 ATPs per peptide bond. So it would cost 131,000 ATPs to make one micrometer of actin, which means you could make 4 microns per second for those energy savings. And so you see that gene loss from the endosymbionts produces so much superfluous energy that eukaryotes were effectively just swamped in ATP. That makes all the difference in the world in terms of what they can do. And this shows that difference. We've already seen this, but this is now measured numbers (Figure 10). This is not theoretical scaling of a sphere. This is the known metabolic rate, the measured metabolic rate, the known genome size, the known polyploidy. And this is a log scale, there are 3 or 4 orders of magnitude difference

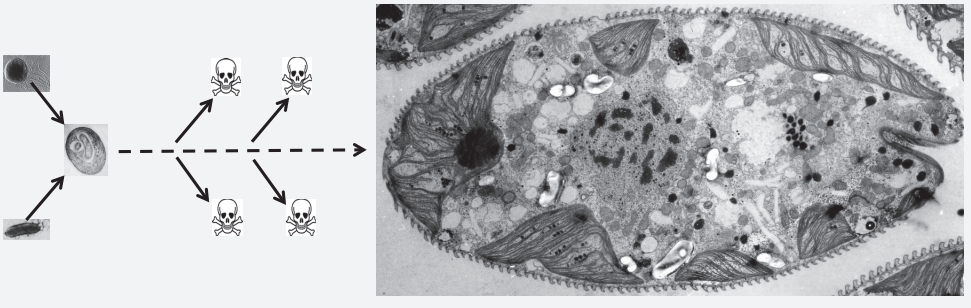

Figure 12. The central problem in eukaryotic evolution. A rare endosymbiosis between two prokaryotes (left) gives rise to an endosymbiotic first common eukaryotic ancestor (FECA). Selection for coadaptation between these cells could arguably have driven the evolution of eukaryotic traits such as the nucleus and sex, ultimately giving rise to a complex last eukaryotic common ancestor (LECA, right, represented visually by the alga Euglena). There are no surviving evolutionary intermediates, implying that eukaryotes evolved relatively rapidly, in small proto-sexual populations. Photomicrograph of Euglena courtesy of M. Farmer.

between bacteria and *Euglena* or large amoeba such as *Amoeba proteus*. We have far more energy availability per gene than bacteria do. Again, there's an overlap between bacteria, archaea and eukaryotes — this is genome size down here, a log scale again, where as up here near the top we see mammals — pretty much anything goes; we can support a genome as large as we want through this method but bacteria and archaea never really get above here. So 3 or 4 orders of magnitude difference again, it comes out exactly the same in that sense.

To finish then, this is the defining signature of eukaryotes and something that we have to wrestle with when we're thinking about treating diseases. Bacteria have a kind of genomic symmetry (Figure 10). Each cell has a similarly sized genome controlling a similar volume of cytoplasm and a similar area of cell membrane[23]. If you were to take a random walk through a population of *E. coli* you would find each cell has a similar size genome. You'd find the same thing if you walk through the cytoplasm of *Thiomargarita* — you'd keep finding similarly sized genomes controlling similar volumes of cytoplasm. But if you were to do that through a population of *Euglena* you'd find a massive nuclear genome supported energetically by these tiny mitochondrial genomes. And so we have a genomic asymmetry. We don't have really a single human genome. There are two human genomes — the nuclear and the mitochondrial genomes. The interactions between the two genomes are tremendously important to human health.

So this is the final slide: why did complex life only arise once? Well it's very difficult to get one cell inside another cell. When we're talking about bacteria, we know of one example, possibly a couple of others that are more equivocal. There must have been thousands of examples over evolutionary time. So it's a bottleneck but it's not a very tight bottleneck. But once these bacteria got inside, we have effectively a simple bacterial cell with other bacterial cells living inside it (Figure 12). This is why we don't see any intermediates — I would say we need to be looking at the interactions between the host cell and the endosymbionts to explain a great deal of eukaryotic complexity[3]. This is the tightest part of the bottleneck — very few cells survive the conflicting demands of living in an intimate union, through adaptations such as sex. We can examine these selective forces by standard population genetics, and we can model some of the outcomes, and make predictions and test those predictions; but I'm not going to talk about that now. I'm just going to stop now and say thank you very much.

References

1. Woese, C., Kandler, O. & Wheelis, M. Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579 (1990).
2. Dodd, S., Papineau, D., Grenne, T., Slack, J., Rittner, M., Pirain, F., O'Neil, J. & Little, C. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* **543**, 60–64 (2017).
3. Lane, N. *The Vital Question: Why is Life the Way it Is?* (Profile Books, London, 2015).
4. Martin, W. Mosaic bacterial chromosomes: A challenge en route to a tree of genomes. *BioEssays* **21**, 99–104 (1999).
5. Koonin, E. The origin and early evolution of eukaryotes in the light phylogenomics. *Genome Biol.* **11**, 209 (2010).
6. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 225–274 (1967).
7. Lane, N. Serial endosymbiosis or singular event at the origin of eukaryotes. *J. Theor. Biol.* **434**, 58–67 (2017).
8. Karnkowska, A., Vacek, V., Zubacova, Z., Treitli, S., Petrzelvcova, R., Eme, L. *et al.* A eukaryote without a mitochondrial genome. *Curr. Biol.* **26**, 1274–1284 (2016).
9. Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature* **191**, 144–148 (1961).
10. Rivera, M. & Lake, J. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76 (1992).
11. Williams, T.A., Foster, P.G., Cox, C.J. & Embley, T.M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).
12. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E. *et al. Asgard archaea* illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
13. Pittis, A.A. & Gabaldon, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
14. Katz, L.A. Changing perspectives on the origin of eukaryotes. *Trends Ecol. Evol.* **13**, 493–497 (1998).
15. Wujek, D.E. Intracellular bacteria in the blue-green-alga *Pleurocapsa minor. Trans. Am. Microsc. Soc.* **98**, 143–145 (1979).
16. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
17. Harold, F.M. *The Vital Force: A Study of Bioenergetics* (Freeman, 1986).
18. Lynch, M. & Marinov, G.K. Membranes, energetics, and evolution across the prokaryote-eukaryote divide. *eLife* **6**, e20437 (2017).
19. Allen, J.F. The CoRR hypothesis for genes in organelles. *J. Theor. Biol.* **434**, 50–57 (2017).
20. Angert, E.R. DNA replication and genomic architecture of very large bacteria. *Annu. Rev. Microbiol.* **66**, 197–212 (2012).

21. Elliott, T.A. & Gregory, T.R. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos. Trans. R. Soc. B* **370**, 20140331 (2015).
22. Lane, N. Bioenergetic constraints on the evolution of complex life. *Cold Spring Harb. Perspect. Biol.* **6**, a015982 (2014).
23. Lane, N. Energetics and genetics across the prokaryote-eukaryote divide. *Biol. Direct* **6**, 35 (2011).