The following content is provided by MIT OpenCourseWare under a Creative Commons license. Additional information about our license and MIT OpenCourseWare, in general, is available at OCW.MIT.edu.

GEORGE CHURCH: OK, so welcome tonight, a very special lecture for me tonight in more ways than I can express. This is the last of my lectures, but I look forward to your lectures coming up. I really do. I've heard from the TFs, and from what I've seen from talking to many of you, this is going to be a really amazing set of projects this year, more than any other year. Anyway, last week, we asked what can biology do for computing?

And this week is going to be a little less nuts and bolts than usual. The entire course-- as you all understand, you're all self-selected to life survey courses, at least this one, because you're still here. And tonight's going to be more of a survey than ever before, and it's going to be talking about topics for which we don't have answers for the most part. There'll be some interesting details along the way. But anyway, instead of what biology can do for computing, this is more about what biology and computing can do for the world.

That's really what we want to do, and also, the higher level network models that we can build since this is network number three. So we'll start out-- we've been talking mainly about cellular models as our highest level of networking, and so we'll talk about multicellular models, in particular, the sort of multicellular models that you get that involve sensory integration and integration of higher levels, and then all the way up to multi-organ systems, where we'll take an example from the atomic scale, all the way up to organ system failure.

And then we'll go from there. So let's start with multicellular models. And this is something that I-- I'll show you a few slides that will remind you of the first lecture. They either were taken from the first lecture-- in this case, this is a wildly different author, and approach, and so forth, but it's the same lesson as from the first lecture, which is we not only have these exponential curves going through the 1900s, but they're super exponential.

They go up, and the trend lines you get depend on which set of points you use, the steepest one being the most recent one, 1995. And this basically is part of when it is that we will have computing power that might be on the order of an integrated human neural net. We talked about neural nets last time as a metaphor, or as an algorithm, but when will we actually have a computer that has the capability of the human cerebral cortex, or entire integrated nervous system?

And the basis of those calculations that Moravec made, and in lecture one Kurzweil was the super exponential curve that I showed. Moravec made them based on his studies on how the retina works.

And I think those of you who have done a lot of computer programming, that understanding how the retina works is probably one of the more intuitive of our various human senses, in the sense that you can program an algorithm that might do in the lower right of slide four here, edge or motion detection.

When a frog sees a fly, black and white, high contrast, zoom across its visual field, something goes off. And you could write an algorithm that could do that, given a few frames of any kind of form. So anyway, but there's been literature on the subject. And when you go through even the best algorithms right now for doing these kind of calculations and you figure out how long it takes to do-- the retina is doing about 10 million detections per second. And then you scale up from that 0.02 gram chunk of retina up to the 1,500 gram human brain, then you get the extrapolation that was in the previous slide.

I think this is interesting both as an introduction to the kind of algorithms we'd like to think about when we think about integrating a multicellular system, but also in terms of where the compute power both biological and computational is going. Do we want to engineer biological systems to keep up with these silicon or whatever other computational system we're using a decade or so from now.

Now, so I said that the visual system is a relatively intuitive computational system to program. In contrast, I would say the olfactory system is less intuitive, at least for me. We know now quite a bit about the molecular biology of the system. There are 1,000 receptors. It's probably one of the largest classes of genomic repeated proteins.

And these receptors are in the same family as the receptors that are the major class of drug receptors, the G protein coupled receptors. And there is basically one receptor per cell, which is quite a trick, right? Because you got one receptor from mom and one from dad. And you've got thousands of receptors. But there's basically one per cell. And they detect olfactory molecule concentrations over about seven logs, where they can detect a concentration threshold with a particular standard deviation.

Now this is the beginnings of the model, just stating the fact that you know about it. When you look at the neuroanatomy, you've got this-- the odorant molecules down at the very bottom of this figure, cilia olfactory neurons, which is a primary signal transduction. This thing goes up where it's integrated in the glomeruli. And you can think of this like the neural nets that we're talking about last time, where you had these interneurons. You had the extra layer that we showed was so important.

And the thing that's amazing about this system are lodged in these four basic olfactory facts, which I take from Hopfield's work, which is really just like this is the same Hopfield that introduced us to or really championed the concept of neural nets as a computational algorithmic metaphor. And here, the idea is you have odor and memory recognition. You have background elimination, just like how your eyes adapt to a mostly red room, your nose is great, or whole olfactory system is great at adapting to background odorants.

So you have one known and one unknown thoroughly mixed. Number three, you can have component separation. You can have a few different odors at once. You have odor separation where you have unknowns. I mean, the combination of these is quite remarkable. So the basis of the model that Hopfield has proposed is this.

You have a coverage for each of the i receptors, where i goes from 1 to 1,000. This will vary depending on the organism. Mice have twice as many as we do, probably, and so on. But the coverage of each of those receptors is equal to the minimum coverage for the concentration of the target, where let's say the target is the target odorant that you're looking for over the concentration that's the threshold for firing that neuron for that target, threshold of t.

And you multiply that times either one or some fraction. It's 1 if you really hit the target. And it's this fraction for that particular receptor and that particular target. You've got some crosstalk. And the amount of crosstalk you can think of this as a field of receptors, which all have variable binding. And that binding is over the six or seven logs here. And this just reflects that same thing that we were talking about.

So the total number of inputs going to the next layer in the neural net, the coverage here, is the sum of this target signal plus the background signal. So the first term is target. And the right hand term is the background. And so you can see they have the same form, the same threshold, and concentration. c sub b is concentration of background. Now, so let's see how this plays out when you actually try this out in an olfactory processing problem.

Here you have an odor space. And what you have here is 80 different neurons is the y-axis at the top of slide seven. And then across the x-axis is time in milliseconds going from 0 to 800 milliseconds. And this is all modeled on fairly realistic parameters based on experiments. And so what you have here is 80 adapting neurons. And what you're doing is two sniffs of a mixed odor. That's what this is modeling, OK?

First, you have a 100 to 500 milliseconds, the range from-- so you can see that something's happening at 100 milliseconds. And up to 500 you have a mixed odor, which is 50 parts of x plus 1,000 parts of y, or 1,000 times y. And then 500 milliseconds you change this ratio very slightly, you see? You just up x by 75 and y up from 1,000 to 1,100.

OK, so that's the paradigm, 50 and 1,000 to 75 and 1,100. And the sniff at 100 milliseconds, that first one, that mixed odorant, it activates more than half the neurons. So this is not like your retina, where if you have a single point source it will activate that one broad cell, or maybe a center surround effect. But it's really activating half of the neurons, OK? And then the changed sniff at 500 milliseconds-- remember, just these two small changes-- is almost invisible, right? So that's what you're seeing right here, OK, is this huge, swamping amount.

But then in B-- so that's what happened in A. B, you've now plotting instead of individual neurons, you're doing the instantaneous rate in hertz, OK? And now, you can see the second sniff here at the 500 milliseconds, which even though you really can't see it in the corresponding point where you're looking at all these individual neurons.

And so you can see that even a 20% spread in one of these parameters is enough to get this kind of easily detectable signal. Now, that's an example of a simulation. That's not a simulation of a particular experimental dataset, but it's based on experimental datasets.

And the code, this is not Mathematica, this is Matlab, which we all wish we could have interchangeably taught in this course. And you can see here in green are the comments. And you can see some of the things that we've been talking about here. The number of receptor types is actually here 2,000 instead of 1,000.

You see the target, remember, I said could be some random number that goes over six logarithms. Well, the way that's done here is here's a random number generated and you generate as the log of the target. And then, you get the target by taking the exponential. It's very straightforward and the code goes on from there.

OK, now this brings us to a very interesting point. Since these three lectures on networks, and in a certain sense, the whole course, is building towards systems biology network models.

And for these models to be useful, in the same sense that the models we have for X-ray crystallography, and the models we've had for genome interpretation in terms of homology, and sequence folding, and so forth needed to be shared. The models, even more than the data, need to be shareable because in a certain sense, even modest manipulation of the data usually involves some good model behind it.

And so these are some of the work groups that are working on different modeling schemes. In particular, I should point out that BioSpice is now DARPA BioSpice, and SBML is kind of a growing part of that. System Biology Markup Language is a way of sharing this data.

And as yet, there's not the kind of convention that we have in crystallography and DNA sequencing where you submit-- at the time that you publish a paper, you submit your data and model, data/model, with an accession number to the database or else you don't get it published. We're nowhere near that now for the rest of biology. But these kind of efforts, probably, are moving in that direction.

And you can see that the features of each of these are some of the things that we've been talking about before. You've got stochastic modeling, kinetic modeling, enzyme receptor cell geometry, neural nets, and so on. This is by no means comprehensive. And there are hypertext links here if you really want to dig into them.

And some of the platforms that they've been thinking about different ways of making it slightly less system-dependent. Everyone has their attempts. Obviously, Windows is not system-independent. So that's two very brief examples of multicellular models, both of them being neural models kind of building from what we were thinking about last time.

But now we're going to a completely different kind of multicellular model, all the way up from individual effects that a single nucleotide-- how much effect a single nucleotide can have on a whole organ system, which in this case will be cardiovascular. So we go from--. And there are a number of physiome and cardiome projects that are of active interest.

Just like the previous slides that I showed showing all the ways of sharing system models, many of these efforts are kind of loose consortium alliances where people have put together different models either to deal with the anatomy, physiology, or sometimes, a integration of molecular to cellular, cellular to, say, neural or muscular, and all the way up to fluid dynamics and so on.

So let's start with a single base-pair change. And I start with this one because you should feel comfortable with it at this point. We mentioned it briefly before. We'll talk about it in more detail now.

This is the single nucleotide polymorphism, which causes the change of the beta subunit of hemoglobin, which is a tetramer, to go from a glutamate, normally, in most people at position six of the beta chain, to a valine.

So that's the purple set of tetramers. Or to tryptophan, and that's the cyan set on the far right. And this, from combinations of X-ray crystallography and three-dimensional modeling, and this is not speculation, just based on primary sequence data, even though the primary sequence here is very, very close, just a single nucleotide and a single amino acid change.

Nevertheless, these authors have been building this from the crystallographic data where not only do you care about the tetramer itself, but how the tetramers interact with one another to make these long, fibrous chains, which are considerably more stable in the sickle cell.

And what you see here is sort of the valine substitution is locked in one kind of conformation and the tryptophan in another. And you can see how these authors have decided how a potential fiber can form in the different cases.

Now those fibers, in one way or another, which is not totally mapped out, but you affect the efficiency of the hemoglobin slightly and the shape of the cell much more radically. And this combination results in the cell, especially under any kind of stress, either oxidative or other metabolic stress, becoming sickled.

So you can have a combination of cells which have varying degrees of sickling. Here in this microscopic slide on the left, you have both the sickle cell and the normal cell side by side. I think when we built up the red blood cell metabolic model, we talked about some of the non-metabolic considerations of that, which were that its function is to transport oxygen and so forth.

And what we're talking about now is more a cell membrane issue. This is on slide 13, where the internal environment, which is the hemoglobin, is greatly affecting now the external environment, which is the hemodynamic flow in the capillaries.

So this falls under the heading now of how is it that we can go from that single nucleotide change to a very dramatic morphological change? In this case, it's pathological, but you could imagine that in the hands of evolutionary adaptation, an organism could take this and run with it. Maybe not sickle cell mutation, but some other one that causes some other morphological change in some other cell or complex aggregate of cells.

Because here you can see a whole variety of different shapes of red blood cells. And remember, these red blood cells are very simple. They have no macromolecular [INAUDIBLE], no DNA, no RNA, so we're really just talking about a bag of proteins which are greatly affected by these different conditions or enzyme deficiencies.

So the system models that were built up, where we had the kinetic parameters for the enzymes, here can, in principle, be modeled on the impact it would have on the osmolarity, or the other membrane properties that were listed in the previous slide, or the sickling that was in two slides back.

So from a single nucleotide polymorphism, we've gone from this three-dimensional fiber of hemoglobin to a change in the three-dimensional structure of this biconcave disk to a sickle disk. But you can also go, sort of in parallel, the same single-nucleotide polymorphism, or some like it, can take you up to pathogen resistance.

We've already mentioned that sickle cell hemoglobin can take you to malarial resistance. But in addition, you can get components of the enzymatic metabolic pathway, such as the one that we modeled a few lectures back, such as glutathione peroxidase. This is part of the redox components.

And here, erythrocytes that are heterozygous for this particular allele should be more efficient in sheltering the cell membrane from irreversible oxidation in binding hemoglobin caused by the oxidant stress that's exerted by the malarial parasite.

And what they observe here is actually an interaction between these two possible haplotypes. So it's actually, you can think of it as a pair of haplotypes. It's the one is the hemoglobin AS, the one we've been talking about for sickling, and then, this glutathione peroxidase.

And so you can imagine, that since both of these things interact with malaria parasite, that your genotype will depend-- I'm sorry, your phenotype in respect to malaria will depend on the alleles that you have at both cases. So you can-- here, you can have an A over S heterozygote and a two over one heterozygote.

And so that's something to take into account when you're trying to do any kind of predictive modeling or modeling to explain the functional genomics that you have in a patient who's a compound heterozygote like that.

So now, the third pathway. Now, from single-- the same set of single nucleotide polymorphisms that affect either hemoglobin or one of the major enzymes in the red blood cell. Now, on to cell morphology. On to pathogen interaction. And finally, to interaction with drugs.

We talked about pharmacogenomics. Here's another example where you have the drug-induced oxidative hemolysis that you occur with certain enzymopathies like glucose-6-phosphate dehydrogenase. This enzyme, glucose-6-phosphate dehydrogenase, interacts with drugs such as primaquine.

And it disrupts mitochondrial function, heme biosynthesis, and so forth, and so on. This is a very significant consideration with a long list of 20-some drugs that has to be taken to account when you have any of a variety of red blood cell enzyme changes.

So we've got these three effects of a limited number of single nucleotide polymorphisms. How do you make this transition from-- and they're all somewhat interconnected. Part of the reason it's resistant to malaria is because it's less effective in its cell shape, and its ability to transport oxygen, and do its metabolism. And same thing with drug sensitivity.

How are these changes in the hemoglobin-- might they be reflected in the three-dimensional shape of the erythrocyte, which is a kind of a membrane-bound compartment? And here's a model that struck me as interesting. It's not been that extensively tested since 1998.

But the idea is that Band3-- this is one of those names that just comes out of the molecular biologist literature, that they ran a gel, and they count the bands, and number three. And it turns out it's about 10% of the red blood cell membrane. It's a very abundant protein.

And it's responsible for the equilibration of anions such as carbonate and chloride across the red blood cell membrane. It's not a pump, it just kind of allows these anions to go across because the pump here is the pumps that are involved in proton and there's also sodium potassium. And that's what your ATP in the red blood cell is going for. And these are just kind of following along.

The idea here is that if you change the degree-- let's see, I think this is covered in more detail. No, sorry.

The mechanism of action here is greatly affected by its disulfide effects. And if you change the redox components in the red blood cell, you get just a slight change in conformation of this molecule.

And then that can result in a net difference between the cross-sectional area of this protein on the outside and the inside, which translates into a net change-- since the phospholipids don't exchange, or the rate constant for the phospholipid exchange is slow and known, then this results in a conformational change.

Anyway, that's the model. That connects the single nucleotide to-- could connect it to the three-dimensional structure of the cell. Then you want to connect the three-dimensional structure of the cell to its ability to carry out its function in the capillaries. And its function there is to allow the fusion of oxygen and carbon dioxide.

So here, each of these cells is shaped by a mechanical process. A mechanical process here, you can take the known or the measurable mechanical properties of a red blood cell and subject them to a method called finite elements analysis where you're solving these partial differentiable equations.

And you can calculate the exchange of the oxygen-- alveolar just refers to-- this would be in the lungs. This would be a small capillary in the lungs. And you can see the capillary now, the endothelial walls are close enough so that the red blood cells are basically deformed in their shape as they go through there.

And the exact shapes that are compatible with this will determine the rate that these little arrows that go from the oxygen on the outside the lung epithelium get through to the blood cell. So that's a reference that deals with that kind of problem.

Now, as we start thinking about building up a larger system model, the other parts of the system, each of them-- so we've got a red blood cell model, it's the metabolic model. In principle, it could also be a shape model and a diffusion model.

But the other parts of the system typically involve muscle cells. They have the smooth muscles throughout the entire arterial and venous system, and most significantly, the heart muscles.

And so here you have an example of action potential in one of the ventricular cells of the dog heart. And it includes all the components that at least we, since this is not an entire course on ventricular heart modeling.

This is the sort of parameters, in terms of all the I to the currents for each of the different potassium channels to K, all of-- each of the ones at I sub K was all potassium. And then there's some in sodium, and sodium and calcium, calcium potassium, sodium alone. And so each of these things, that internal storage of calcium [INAUDIBLE] from one subcellular compartment to another and so on. You get the idea.

Each of these things has to have the parameters measured. And any of that are absent, you have to have reasonable ways of getting surrogates. And so this is just the sort of data that would go into the other major type of cell that comes into this cardiovascular modeling.

And finally, you can integrate this up to a fairly complete system. It's obviously not complete until you get to the whole organism, but higher level, and how we're talking about whole body recirculation. This NSR group, this is a hypertext link for this particular model which you can download.

And it has a four-chambered heart. We were just talking about one ventricular cell on the previous slide. But that would be-- if we look at the second box from the top here, this is where the heart and lung model would fit in there. It's got seven different organs, which include kidney, liver, lower limbs, and so forth down at the bottom part of this diagram. And you get the picture. Each of these things, you're modeling the volumes and flows,

Now, this is systems biology. This is really kind of a renaissance of interest in physiology, which is what it would have been called before. And there are some really interesting mysteries out there to be solved that are really only solvable, or possibly only solvable, at the system level. You never know, somebody could come up and say, oh, yeah, this is really well understood immunology and that answers everything.

But even if it's some kind of immunology, you still have to say, how does it play out? And what happens is we've been talking about sickle cell disease, and you can have fairly mild pain as one of the major symptoms. But then all of a sudden, out of nowhere, you'll get this multi-organ system failure where almost every major organ in that previous slide, lungs and so forth, fill up with fluid and you have a very high chance of dying.

And this could happen to every one of us in this room because it's not restricted to sickle cell. If any of you have severe burn, or car crash, major bone injuries, and so forth, you have a very good chance of getting into multi-organ system failure. And it really isn't known how that plays out. I mean, not even that-- not just a quantitative model, but even qualitatively.

So I think this is something-- there are projects now to get genomic data collection, both genetic and expression data, and time will tell whether that is actually the best route to solving that mystery. So that was multi-organ. Now we're talking about multi-organism.

After you've built up an entire organism, then you want to know how it acts at the next level of network analysis, which is how does it fit in with other organisms? None of us, almost no organism, really belongs in ecological niche all by itself. And some of the modeling we are taught in this course could be called simulation.

And probably one of the longest-lasting computer games in the world, and one of the most successful, I've heard, is *The Sims.* It started with *SimCity* in-- *SimCity* in '87. And then this particular one illustrated here is *SimLife,* which is all about ecological modeling.

And it's not entirely different from what a serious ecological-modeling program would do. It certainly has a lot of the interesting parameters such as the lifespan here. You're doing demographics here so you have the lifespan, the amount of food needed, the size, kind of vision, roaming, so forth, of all these different kinds of animals.

You can have plants, you have herbivores, carnivores, and so on. And then you can set up the population sizes, and then it will do simulations based on that, where each individual animal is tracked in terms of its position and quantity. And as you would expect, as the carnivores build up, then they knock down the herbivores, and the plants go up, and so forth. And it goes through these cycles.

And so this is basically a stochastic model, just like the stochastic models that we had for molecules, but here, at the organism level. And what's happening here in the slower right-hand section is you've got little green plants getting eaten by herbivores, and the herbivores getting eaten by carnivores.

Now, hopefully, this course has already or will inspire you to really think globally. To think not only globally in terms of how little pieces of molecular tools fit together into systems and networks, but how our systems and networks we model fit into the big picture of what are really important problems.

Maybe a slightly improved Viagra is not in the same category as a new tuberculosis drug or maybe it is. I mean, you decide for yourself what is thinking globally. But you have to act locally, and that's what we're doing in this class.

When we're thinking globally, and we're thinking about ecological systems, we're thinking about the lithosphere, for the most part, and its interaction with the hydrosphere. The lithosphere is mostly silicon dioxide, a tiny bit of carbon, and it gets very hot, very quickly.

So only the top 0.1%, that's the top four kilometers of it, is survivable in any type of organism we know of. About 110 degrees centigrade is where organisms start having a great deal of trouble. You and I would have trouble before that.

The biosphere here is about 3 times 10 to the 15th grams, my estimate for marine organisms. And maybe 10 to the 18th grams for all land plants and animals and microorganisms.

The microbial hydrosphere here is about 10 to the 21st milliliters, which works out about 10 to the 27th cells. A phenomenal number of these cells, about 10 to the 26th of those cells is a single species, which is prochlorococcus, which is responsible for maybe about 50% of the Earth's photosynthesis.

A lot of that photosynthesis does not end up in fixed carbon because it is immediately consumed by one of its other predators, basically. But there's really quite a lot of cells out there which, to the extent that they're well-mixed, and they're not-- of course, it's not perfectly mixed, all the Pacific and Atlantic and so forth.

But it's a considerably more well-mixed, say, than the lithosphere organisms, where you have organisms down a kilometer into the ground. They don't move around very rapidly. So when you have a population of that size as you remember from one of our first lectures on population size, the effect of population size determines the rate of drift and the optimality of the organism.

So one of the things that we do, when we've been talking about mining the biosphere, one of the things that we're looking for are new tools that we can use for nanoengineering.

And kind of one of my pet ones that I'm becoming interested in, we don't work on the same thing, but it is kind of interesting, is a set of-- we mentioned this briefly in the drug protein interaction lecture where-- but I don't think I mentioned this particular one-- where you have polyketides that go together.

It's another polymer that has certain similarities to the basic polymers we talked about, but each step in it had to have a protein enzyme-- an enzymatic domain to accomplish that. And so one of these is tetracycline. And this is one of the more aromatic kind of coupled-aromatic compounds that's made because you have these [INAUDIBLE] and aromatases in addition to the polymerization steps.

And so you make this thing that looks kind of like a poly aromatic hydrocarbon. And also in nature, you will find, in soot, in forest fires, in all kinds of natural phenomena, you will find polycyclic aromatic hydrocarbons, which some of you may know about as the potent carcinogens, but they're also just natural components.

But they also-- you can start to see this looks a little bit like the buckyballs and buckytubes that we talked about when we were talking about molecular-type transistors. So the possibility of mining the biosphere for enzymes that act to synthesize or degrade this class of compounds, I think, would be just an example. We could list many others.

But the part of it is just to dream, to imagine, what it is would like to find out there. And if there's an abundant source of it, then you should be able to-- abundant source of, say, the compound, in this case, you need to be able to find a microorganism and an enzyme that goes with that because there's a truly phenomenal amount of diversity.

Another very important global consideration, rather than just mining it for new tools, is thinking about ways that either we could engineer or accidentally mess up our entire planet, as we could be doing with global warming or could do.

And perhaps it's naive to think that we actually are having such a big effect because we know that global warming changes periodically over millennia. But it is very clear from the record just exactly how much carbon dioxide we are releasing and it makes sense, that it is consistent with the kind of temperature changes that have been observed since the Industrial Revolution.

In any case, when you look at, in particular, the Southern Ocean-- up at the top of slide 27-- you can see there is a few places, in particular, Southern Ocean seems like it's a prime candidate where you have high nutrients but low chlorophyll. And you say, well, why would you have high nutrients but very little of the chlorophyll around, which is a tipoff that you have the photosynthetic bacteria.

And the reason is that you have a limitation of some micronutrient. Micronutrient means it's not needed in the vast quantities that you need nitrogen, phosphorus, carbon, oxygen, and so on. So iron, typically, is the limiting micronutrient in the Southern Ocean.

And so there actually have been little pilot experiments to drop iron off the backs of huge tankers. And there are now at least seven patents that have been filed on doing this in order to balance out the carbon credits for different nations. This is potentially a very big bit of terrestrial engineering that might happen. It involves trying to change phytoplankton.

And as this line here begins to point out, even though phytoplankton are only 1% of the total global biomass, it's about 50% of the carbon fixation. What is the source of this 50-fold anomaly? Why is it that even though it's doing-- how is it doing 50% of the carbon fixation?

And what's happening is a lot of the carbon is going right back out after being fixed. Instead of settling to the bottom of the ocean and never bothering us again, it gets returned. And the exact modeling of that requires a much deeper knowledge of exactly what organisms are present in the ocean.

One of the problems with this however, is a genomic problem, in a certain sense, which is that very few of these organisms can be cultivated in the laboratory. On the order of 99.9% of the organisms that you sample from the ocean, or from the soil, a number of different environments, do not grow well in the laboratory.

So if you were to study them, some people are feeling that the best route to studying them is going directly for-- looking at their genomes, without necessarily being able to grow them in the laboratory. Now, when we talk about this problem of-- even if we fertilize the oceans, and we can model this whole procedure, I won't go into details.

But if there are predators there that take the fixed carbon and turn it back into carbon dioxide, you want to be able to monitor that. Now, of course, the ocean is a large, complex set of prokaryotic autotrophs that are, say, photosynthetic bacteria, and eukaryotic photosynthetic bacteria, and prokaryotic and eukaryotic predators of various sorts.

And this is, I think, one of the more interesting predator-prey differential equation models. It's almost the same as many of the other differential equations that we did from the first lecture on growth. We did the logistic equation, some of the ones we did on repressor function, and so on.

But the thing that's kind of interesting about this one is it's not simply a single-cell bacteria being eaten by a single-celled heterotroph, let's say, a blue-green algae being eaten by a single-cell heterotroph. Here, they are both multicellular. They're sort of the minimal multicellular.

But because they are-- so Chlorella is one of the smallest plants, multicellular plants. And Brachionus is a small rotifer, a multicellular animal. Because they're multi-cellular, though, now you have the demographic mortality and fecundity of each of these things that has to be modeled in. That's the m and the lambda for Brachionus.

And so this plays into these equations where you have here, in the upper-right of slide 28, the rate of change of nitrogen with respect to time, and concentration of Chlorella. So nitrogen is N, Chlorella is C.

The R is the Brachionus that are reproducing and B are the Brachionus which are the total. And you can see each of these equations and how it plays out. Now this is how-- and the way they play it out here is you do it by dilution rate, delta, here.

And the dilution rate is shown along the horizontal x-axis for each of these three plots here. And what they're modeling is on the far left, the nitrogen concentration as a function of dilution rate. And you can see you get these different sectors of behavior.

When you look at the Chlorella, this is a green, photosynthetic, multicellular organism, in units of millions of cells per milliliter, that can take different-- it take different pathways here where, as a function of dilution rate, you get this bifurcation where you get a huge split in the possible concentrations of Chlorella is the one curve and the Brachionus, the predator, is the other curve.

This kind of bifurcation is the sort of thing we talked about earlier, both in the logistic equation and in the self-exclusion modeling. And in the upper-right-hand, here's our friend the coefficient of variation in percent, ranging from 0 to 80 or so on the vertical axis and the dilution rate.

And again, you can see that you get a peak in the dilution rate at around 0.75. The bottom two are models, and the upper right is the data. You can actually see, when you run real Chlorella and Brachionus in here, that you do get a peak just as you would predict in this kind of bifurcation analysis in kinetic modeling.

That's due to-- where you get interesting behavior in this complicated-- or this fairly simple model ecosystem where you just have two species. In the ocean, of course, you have a lot more. Now, that's out in the wide world where you have 10 to the 26 cells.

But inside every one of us in this room, no offense, but there are about 10 times as many non-human cells in each of us as are human cells. And we have our own ecology. And there is a literature where certain organisms we know cause diseases, and they're put into one category, infectious diseases, but then there's another set that takes some time to sort out.

And you can see here are candidate diseases and candidate organisms. And all these little bars that are going in here are references that you can look up where maybe it's not an airtight case yet, or maybe it's not even-- maybe there's controversial or discredited [INAUDIBLE], but these are links where, eventually, this will be called an infectious disease.

Helicobacter pylori causes stomach ulcers. I think that fact's basically already accepted. Some Australian doctor named Marshall decided to drink a bunch of Helicobacter to prove to his colleagues, who didn't believe any stomach ulcers were caused by Helicobacter, much less all of them. And he drank it, he cured himself, he drank it again. And he caused stomach ulcers in himself.

And hopefully, none of you will volunteer to do similar things with some of these other nasty guys here because I think some of the list of things that they cause are a little more serious than stomach ulcers, and the cures are a little less obvious than the cures for a gram-negative bacterium like Helicobacter.

There have been various efforts started to actually mine what's the connection with genomics and computational biology? To mine the transcriptome for evidence of these. You can look for bacteria very easily because they have ribosomal RNA components which you can PCR.

Viruses are more complicated because they're not conserved, or they don't have any universally conserved element like ribosomal RNA. But there have been efforts to look for what's present in the human transcriptome that's not present in the human genome. The human genome is something highly purified and cloned and so forth, while the transcriptome is whatever cells they ground up that day and they.

And indeed, lurking in there are many of these hepatitis, papilloma, Epstein-Barr virus, retroviral-like elements which are not present in the human genome as yet. Human genome is not completely finished. So that's still an escape clause.

But basically, these are smoking guns for at least commensals. Microorganisms and viruses that are living in tissues, some of them may be tissue-specific. Some of them, subset, may actually cause disease. And that's the whole problem then, is sorting out cause and effect. We know how it's been done, heroically, with Helicobacter pylori, but how do we do it with the other ones? Some of them will have tissue-culture models.

At the time this slide was done, but so-- this is just illustrating kind of the flow of nucleic acid information that comes in about-- that was capable of being mined for new microorganisms, new viruses that might be present in a tissue-specific fashion, might be pathogens.

These are some of the most popular sequences from which we're mining. Obviously, human was popular at times [INAUDIBLE] the human genome was in place. Here's Brachionus, our friend, which was pretty high up on this list, considering most of you probably didn't hear about Brachionus before this lecture. Now you've heard of it twice. There just was a spurt in sequencing interest.

And of course, HIV is the winner year after year because of the importance of resequencing it for new mutations. And it really has a record number of new mutations. And we've talked about HIV from a variety of different standpoints. One of them is polymerase, is protease, is drug targets, and as a source of drug resistance.

And that you can follow up that model, that sort of atomic-scale protein modeling, where you're looking for new drugs where the polymerase, the mutant polymerase, will not be resistant to the new drug. But you can also monitor at the population scale. As HIV goes to a particular patient, or to a population of patients, how does the drug resistance change as a function of time?

Here, the horizontal axis is in days, up to 30 days, and then the vertical axis is the titers of the viruses [INAUDIBLE] as a function of time. And what you have here is rates of exponential increase. These are all on logarithmic scales.

And where you're modeling such a thing as the clearance to the immune system and so forth. Originally, this virus was thought to be a very slowly replicating virus, almost cryptic. And then later was found out this is actually a very rapidly replicating, but immune system is very rapidly responding.

Just each of these models we've gone through has a set of parameters that it goes along with. In a certain sense, there's quite a bit you can learn about a model just by seeing what parameters are present or absent, whether they're experimentally determined, how accurately, and so on.

And model parameters here are mortality rate of uninfected CD4+ T-cells, same thing for infected cells, and they can have different [INAUDIBLE] you can see on the far right that there's a huge difference between a diet of 1/4 per day versus [INAUDIBLE] and so on. There's rates for getting infected, rates of virus loss, production, threshold value for remission, so on and so forth.

And these are important parameters for how the virus population changes within a person or the population. How it spreads through the population depends on, in many cases, on the herd immunity and so forth. And we get into issues of public health.

It's almost a truism that most of the additional quality years of life, quality-adjusted life years here, QALY, that we have in the world the fact that our life expectancy is so much longer and so forth, are mainly public health decisions that have been made, not so much pharmaceutical.

Such things as clean water have made a huge impact. And so we need to think-- but even when you think about the pharmaceutical aspects, it needs to be thought of in a public health sense. And a surprising amount of public health officials in even the most developed nations, maybe most of them do not have formal education in public health. And so I urge you all to get at least some education in this.

Because in this, we have-- here, they're trying to build and have built quantitative models, not only for the kinetics which a disease will go through a population or even through an individual, but the way that you make decisions. And this is having an impact on the way decisions are-- how these different projects are prioritized.

At any given time, you might have hundreds of different candidate vaccines, which is one of the major, very effective public health strategies. And this is the way they prioritize it. So the level one, saves money and it improves life. So that's almost a no-brainer.

Then you have different levels. We have it might cost a $10,000, or $100,000 per quality-adjusted life year saved. And so the level one candidates are cytomegalovirus, and therapeutic vaccines.

These are not aimed at an infectious disease, they're for diabetes, rheumatoid arthritis, multiple sclerosis, various bacterial ones. And of course, HIV was not even in this study because it was such a high priority already within the NIH.

So is there a role for genomics and computational biology, the title of this course, in this vaccine research and development? I think the answer is yes, but it requires a great deal of creativity and resourcefulness on all of your parts. There are new opportunities with DNA vaccines where you can have one or more DNAs shot intramuscularly or a variety of other ways.

What is delivered by the DNA vaccine can be various so-called intracellular vaccines, which can either act through cell-mediated immunity or in some other way intracellularly. RNAi is a rapidly emerging way of delivering things that may not be classically considered therapeutics or vaccines.

The concept of multiplexing-- many of us have received multiple vaccines at a time. Typically, every year, your flu vaccine will have two or three strains in it.

But as you can see from this middle article, the diversity of certain diseases, like HIV and influenza, almost demand attention from genomics. The diversity that can occur, either from year to year or throughout the entire global set of viruses.

Another opportunity is when we have vectors for, say, arthropod vectors, insect vectors like mosquito. And we have opportunities to hit malaria and all the different life stages. There are multiple life stages within humans and multiple stages within the insect vector.

And now, the genomics of these two just came out recently, in the same week, in*Nature* and *Science.* This provides a whole new set of inspirations for work that can be done on these. I think we should take a little break now, and we'll wrap up the talk after that break.