GEORGE CHURCH: OK, welcome back. We just finished our discussion of predator, prey, and host parasites illustrating it with ways that we can have an impact on ecological modeling in oceans and in public health. And these kind of considerations start getting us into what are the global and socioeconomic considerations? What kind of impact do these kind of models have on how we make decisions?

And in this context, I'm glad to be associated with the Genome Project. It's one of the first scientific projects of any reasonable scale that had from the very start, from the very first proposed funding for it in 1990, a component of 3% set aside for ethical, legal, and social issue or LC. And some of the conflicts that are covered by the grantees in this part of the Genome Project were genetic non-discrimination, privacy, reproductive rights, cloning, psychological stigmatization that can come from maybe too much knowledge, clinical quality control, what can happen with false positives and false negatives in clinical exams, safety and environmental issues, such as the ones we were just talking about and some more that I'll raise.

Uncertainties, not just in quality control, but in the uncertainties in testing minors. Issues of diversity, both biodiversity and human diversity, and commercialization of the products. Who owns my cells when I give them to a hospital or a company to help cure me? Do they then own all the patent rights? So the underlined topics are the ones, the four underlined topics we'll talk about in the next few slides. In terms of non-discrimination, this can go any direction. It will go the direction the market forces push it, meaning the voters.

If we want to pass really laws that say that if George Church sequences his genome and he has to report that sequence to his insurance company or worse yet the insurance company is going and get it before he wants it or not. But on the other end of the spectrum, and I hope the trend which will make it much easier for us to share our data, or use our data-- that's it-- is that the Non-Discrimination in Health Insurance and Employment Act was [? docked ?] in 1999, was introduced and passed, which would extend employment protections in the government sector to the private sector to the extent that that can be generalized.

I don't know exactly where that's going but one can hope that-- I mean clearly we will always do some kind of discrimination based on genetics which can be assessed such as in an interview, you know, height, friendliness, that sort of thing. But we won't necessarily be doing it based on genetic sequence. And these are tough issues. It's the probabilistic nature of the decisions that you make during an interview different from the probabilistic decision you would make based on a DNA scan? Which is more accurate? Do you want it to be more accurate? Do you want it to be less accurate? So on and so forth.

But clearly there's the trend in terms of legislation is towards less information being used for discrimination. Certainly in the insurance level. And it's appropriate in insurance because in a way that's supposed to be a process by which risk is shared. So the issue of races. Typically races, some people feel this is a scientifically rigorous definition, others do not. Some feel that it involves very, very broad strokes in other organisms at least. It can cover fairly detailed bottlenecks in populations. But in any case, there are elements of population structure which certainly are important, whether you ascribe it to the major races or to minor, smaller populations.

And examples here, we've already talked about hemoglobin variants evolved to resist malaria. It's going to be one of the themes today. And as with the differences in skin pigmentation, the pressure of the environment to develop a group-wide trait was powerful and can involve a very small number of genes. You get founder effects where a particular population is highly enriched for a particular disease such as Huntington's in the case of Lake Maracaibo in Venezuela or Tay-Sachs in some Jewish populations.

So when you have these well isolated populations, I think, you can see in the population genetic literature how they can be used in various ways. But, overall, and it's sort of the broadest level we all share a large set of commonly acquired polymorphisms. And even with ants, the smallest subpopulation, it still behooves us to look at the details of the individual variations in DNA sequence in the haplotypes or genotypes.

I have two dangerous slides in a row. I could probably come up with quite a few more, so could you. But one of them is-- this is really a very strong advocate for her modeling. And it's totally anecdotal slide itself is that is non-scientific, historical. But this is one critics' view of the huge influence that Lysenko had on genetics and biology in the Soviet Union from the early days of the revolution.

And his real pet theory was the inheritance of acquired characteristics. And there's nothing intrinsically wrong with in broad strokes the inheritance of acquired characteristics. It's not certainly not common biological phenomenon. And at the time, it was not helping their biology progress at the same rate as the rest of the world which was mainly pursuing the Mendelian inheritance models.

And this critic in quotes felt that his habit was to report only successes. This is a really feel good habit. And his results were based on extremely small sample sizes, inaccurate records, and the almost total absence of control groups. He made an early mistake in the calculation which caused comments among other specialists in his field and made him extremely negative towards the use of mathematics and science.

Making mistakes should not cause you to drop the use of mathematics and science. Hopefully, those of you, probably everybody in this class has made a mistake in mathematics. And, hopefully, none of you will drop mathematics in your future biology. And the second danger is the danger of ethics-free science. Now you get to the truly remarkable story that in 1979, there was a release of anthrax 836 spores and part of the former Soviet Union. And a few years later, actually 1999, this book was published describing what was behind that.

And what was behind that was decades earlier in 1953. There was a leak in one, I guess, of their anthrax developments. And then 1956, they found that one of the rodents that they captured in their routine surveys of the sewers searching down a possible anthrax, that actually a strain had become much more virulent than the original that they were working with.

And usually the response of public health official or even a reasonable person would be to kill this thing. But instead they decided that this was great. Let's cultivate it. And the idea was to install it into these rockets that were targeted on Western cities. And then that led eventually to them depositing spores on their own people accidentally of this greatly enhanced strain.

So the question to our community as this genome engineering tools get easier and easier where you basically can sequence genomes inexpensively and even synthesize genomes inexpensively. And much of this is going to be in the public domain, what is to stop this from being a very easily disguised and potentially very inexpensive form of terrorism?

And part of that is that we do what we can do to improve either detection tools or genome engineering tools to make them more of a defense type than offense. But this was done in the early days of recombinant DNA where the vectors were designed such that if the vector ever were to escape from the lab, it would die immediately. It would be lacking the nutrients it needed to grow. It would be very sensitive to detergents that occur typically in sewers and so forth.

Now the vectors we use today, those were not very robust. They didn't grow very well in the lab as much less in the sewers. But it turns out that random release is fairly low risk problem. That was what we were worried about in the 1970s with recombinant DNA. But purposeful release of genetically modified organisms is more of the issue today. And, of course, genetically modified organisms on this slide have been created and engineered over the millennium, maybe 10 millennium or so, without a license.

And they've been very successful. You picture this little weed-like thing at the top left, and you end up with this great 4th of July corn on the cob. And it's a hybrid corn. And similarly, these dogs range over three logarithms in adult mass would barely be recognized as the same species if we didn't know them and love them very well. But these are examples of genetic engineering that was done pre-genomics.

But now when we use recombinant DNA in particular, but genetic any kind of interspecific genetic modification, this definitely raises environmental issues, especially in the developing-- I'm sorry, in the say European nations where they are wealthy, and they don't need much improvement in their agricultural needs.

Some of the developing world has very definite needs for genetics, or feel they do, for genetically modified organisms. Some of these include producing vaccines in plants. Vaccines are one of the most cost effective ways of generating public health results but possibly even more effective would be to have your bananas and other crops contain vaccines. And these have actually been developed. But it's been hard getting them actually supplied due to concerns about release of genetically modified organisms. Possible allergic response and so forth.

Salt and drought tolerance is extremely important. They often come together. And there are a huge number of drought and salt tolerant plants, so-called resurrection plants. About 100 different, completely different species which could provide new genes that can be introduced into non-drought tolerant plants and are being introduced by scientists in developing worlds, such as Africa.

Terminators were originally, it's like this popular, unpopular, popular again. They were popular with the companies that developed them for reasons that we may not know. But there was the concern or the backlash was the companies are doing this to preventing the farmers from reseeding. By the terminator meant that the next generation seeds would be useless. And it had been the habit of a farmer to reseed. And so the terminators didn't sell for a while. But then because of the worry about dispersal of genetically modified organisms, now terminators are becoming more interesting again to a broad set of people.

You talk about organic farming, there is controversy as to whether the tightest definition of this, which involves no inorganic fertilizers like nitrates and phosphates and so forth. But that means a very high animal load. High animal load means you need to use up a lot of your vegetable crops in order to produce the fertilizer for the other vegetables.

And as you pointed out that natural is not necessarily harmless. A wide variety of naturally occurring compounds, natural pesticides, are also carcinogens. And here's the laundry list at the bottom of slide 43. Cloning of stem cells. We have problems with cloning definitely. In almost every species, even the successful ones, there are a variety of species which have been tried and not succeeded. There's some that have been tried. And if you look at the studies in detail they're examples of developmental defects higher than expected in that species ranging from a few percent on up.

And, obviously, there's some kind of epigenetic reprogramming that's occurring here where you're not getting all the right contributions from the maternal and paternal genomes. And this can be possibly studied with expression profiles. We can start employing all of our automation to analyze how we can increase the fraction of stem cells that we could take adult cells and send them into different lineages, or we can formulate ways to either transform one adult stem cell into another or an adult stem cell into a slightly more primitive stem cell and still retain the advantages of so-called therapeutic cloning, which would be to maintain good histocompatibility, say, with the patients.

So finally, education. Why should we bring up education in a course? We've talked about models of decision making in public health. But there's also a similar set for education. We want to be able to deal with uncertainty, complexity, quantification. I'm sure that I have introduced, or we, have introduced plenty of uncertainty and complexity in quantitation in your lives with this course. I apologize for the parts that are painful. I hope that no pain, no gain has some applicability here.

We want to-- a theme in this course has been to cherish your exceptions. Collect them, and these can be discoveries. They are at least going to keep you honest and keep you from making big mistakes. We want to be able to translate from one data type to another, from one conceptual foundation to another, to integrate different either adjacent conceptual spheres or very distant ones. The way we do this, slide 47, we need to have measures of our progress, measures of the quality of the underlying data and the models.

And we've already done this to some extent, basically, for three dimensional structures and sequence data, the primary and tertiary information. For X-ray diffraction-- this dates back into the '60s-- we have measures of data quality that include resolution, model quality, which is the R factor we talked about before. We have ways of doing similarity searches.

For sequencing, a more recent enterprise in the '80s, late '80s, early '90s, we got the Genome Project launched with thoughts of measuring data quality in terms of discrepancy per base pair. That should be less than 1 part in 10 to the fourth. The models, typically, are models of protein conservation. And a similarity search is one of the greatest killer applications of all times, which is [? last. ?]

And then for function we're less far along. We don't really have accession numbers, as I said. We don't really have great ways of doing similarity searches through say image databases or similarity searches other than say correlation coefficients. But I think this is rapidly changing. When we start applying these models, this is not only in the education sense, but in the probing these networks that we've been talking about, even further beyond consideration of neural networks and our interaction with other organisms, kind of combining those two, the neural networks and our interaction with other organisms, ecology, is this notion of biophilia.

Which is that we, as human beings and other fairly intelligent animals, are connected subconsciously to other living beings. It is clear that no matter how urban or what culture you're talking about, snake dreams figure in very prominently. And this has to do with the need of primates to avoid and track snakes from their vicinity and the vicinity of the tribe. Little animals are cute. We just know that. And they are much cuter than the adults, no matter what the animal is, with the possible exception of snakes.

[LAUGHTER]

And there are anecdotal, at least, possibly well-characterized effects of the green fractals that we like to see in trees and plants. And taking this one step further from just this kind of very stimulating thought that we actually have co-evolved, our nervous system has co-evolved for so long with these other things. It would be natural for us to be very attuned to them because our survival would have depended upon our ability to avoid snakes and to deal with green versus brown shrubs.

And if you take that one step further, much of what we do in the humanities say is affected our aesthetics and our approaches to it, our beliefs are affected by this heritage. And this is what E.O Wilson has championed. And I would say it's quite controversial as it should be.

But in general, long separated fields come together-- it's this consilience definition-- and they create new insights. Like chemistry and genetics brings us molecular biology. And the question is, is all of human endeavor ready for such a thing, from religious feelings to financial markets and so on?

And whether or not, how might genomics and computational biology contribute? It's surprising sometimes when these things do contribute. And here's some speculation on that. We have improving imaging methods. We talked about imaging in the context of in situs. Some very functional versions of this are positron emission tomography and magnetic resonance imaging, which allow you to monitor such things as blood flow or metabolic uptake in different parts of an actively metabolizing brain.

And here you can inject some O-15 labeled compound. And you can get an image resolution on the order of nine millimeters. OK so the voxels here, volume elements, are limited in that range. And sort of in the 20 seconds is sort of the time frame that you're working in. And this has been applied to a whole variety of interesting behavioral tasks. Almost anything you can think of counting memory, so forth in humans.

And you can map the parts of the brain that are differentially responsive to a control and an experimental time frame for the same patient in the same apparatus. And here, this is just to show you how far I can go and connect to what we were talking about in the previous slide, here religious subjects, patients, have been the differential parts of the brain monitored here are in a religious recitation process and in arresting control. And these are all the P less than P to the 0.001 significance by the very complicated statistics that are used in the analysis of these functional maps.

Now not to leave us on that note but to connect it more to genomics and computational biology is more in the heart of this course. Here's an example how magnetic resonance imaging, which can be applied to gene expression, how we bring these two together, kind of imaging gene expression. And one of the powers of this in contrast the positron emission tomography had about a 9 millimeter resolution. This has a 10 micron resolution, sort of in the optical resolution range. But unlike optical methods, this will work for impact opaque organisms.

And, finally, so MRI will do that in general. But to connect it to gene expression, you need to tag the gene expression such way, and we have green fluorescent protein, we have other colors fluorescent protein. We have luciferase, data glycosides, and so forth. One of those has to be turn on something that has sufficient contrast and magnetic resonance imaging.

And an example of that is a way of caging a gadolinium ion such that you can see this little galactic tyrannous ring at the top right is covering it. And the little red bond there is cleaved by beta galactosidase, lacZ fusions. And so you can now make your same reporter constructs that have been used for making, say, blue cells with a core metric indicator with optical wavelengths. Well, now with magnetic resonance, imaging released that gadolinium. So it's now accessible to the solvent, has a different resonance, and you can get sharp contrast here.

In a living organism on the top, with 10 micron resolution, as opposed to a fixed organism, which is required for getting the whole-mount in situ staining with lacZ. So I think, this is a very interesting combination of being able to use something alive in opaque tissue with a genomic tag of sorts.

Well, on that religious note, we will sort of wrap up the course, my part of it, the lecture part. And in so doing, I want to especially give many, many thanks not just to this year's TFs, which we will get to in just a moment, but the ones that helped start this phase of the course. Actually, the course goes back to '88. But these are some of the Teaching Fellows in 1999 and 2000, 2001.

One of these Teaching Fellows, [? Suzanne ?] [? Camille, ?] has stayed with us as a head Teaching Fellow this year. She was up to the top of-- and I'm very thankful to every one here, Woody, [? June, ?] [? Len, ?] Tom, [? John, ?] [? Juang ?] [? Hong, ?] Gary, and [? Lachmans. ?] If any of you are here, could you kind of stand up, wave, just wave. Thank you very much. Really, I love it. And if this course is to continue to survive, we need TFs for next year. So if any of you feel that you have the right stuff and feel that this course should survive, please contact us.

Hopefully, it'll be some finite positive number. We really love the projects, and we'd love-- and in the past, many students have been reluctant to stop working on the project after the course is over and their grade has been assigned and so forth. And that nothing could make me happier than to have do that. There's a limit to just how much I can help you do that.

But if anything I can do to help, providing additional mentoring and so forth, I'd love to do that. Some of these have even resulted in publications later on. We started this course, slide four of the first lecture, actually the first real slide or the first lecture, was on the origins of zeros and ones. Kind of a play on the 101 course number, where did the binary code come from? And you could ascribe it to Leibniz or one of the modern 18th century mathematicians.

But Leibniz himself found that it had already been invented about 5,000 years ago or so by China's first emperor. And since then, some people have gone so far as to take this binary coding, the etching, which has 64 hexagons and has arranged it in such a way that it has decoded it so that it actually fits in with the geniculate code which has 64 codons.

So for those of you who are students of the etching, hopefully, it now has new meaning for you. And to emphasize this yin and yang symbol here, remember that the purpose of having the black dot in the white zone and the reciprocal white dot in the black zone is to remind you that things are not just black and white. They're not just zeros and ones. They have more complexity, and they are constantly in change. This is the book of changes, and this course is about change.

And you should-- you do computational biology. It's not just about computation. It's not by keeping yourself busy at the computer, keeping your computers busy all the time. It's about thinking, thinking as broadly as you can. So thank you very much.

[APPLAUSE]