So far, there are a lot of ways to clone genes. I've given you two. Later in the course, I'll give you a third, but we cloned it here by its function, complementation. We cloned it here on the basis of the protein it actually made. Now, how do we analyze these genes? How do you know which gene we're dealing with? How do we study it? Let's turn to that. Analyzing our gene.

Well, let's suppose we have our cell, maybe it's our yeast cell. It's Arg1 again. We'll go back to yeast. It's got its yeast DNA it's also got this circle that has our plasmid here. Circle is plasmid. How are we going to study that gene? We know that yeast is able to now grow. We can grow up that yeast. We can purify DNA from it.

But wait a second. How do we get our plasmids away from the chromosomal DNA? Aren't we back in the same boat of you can't purify DNA from DNA? Thankfully, this is easy. Little circles of DNA have different biochemical properties than the long, linear chromosomes or the big chunks of DNA. And you can use biochemistry to separate little circles of DNA from big chunks of DNA.

So no problem. You can purify the plasmids. Purify the plasmid, and there it is. It's back again. And now, you've got your Arg1 gene. How do I study it? Well, the first thing I might want to find out about my Arg1 gene, is how big is that insert. How can I find out how big that insert is?

**AUDIENCE:** Electrophoresis.

**ERIC LANDER:** Electrophoresis. I first have to cut out the insert. How do I do that?

**AUDIENCE:** The restriction enzyme.

**ERIC LANDER:** My restriction enzyme, EcoRI. I cut with EcoRI and it liberates the insert. And then I take a gel, which can be made out of different substances, and I put my DNA there. Agarose is particularly nice stuff, which is basically Jell-O. And I turn on an electric field. DNA is what, positively charged or negatively charged?

**AUDIENCE:** Negative.

**ERIC LANDER:** So what kind of charge do I want to put here?

**AUDIENCE:** Positive.

**ERIC LANDER:** What charge do I put here?

**AUDIENCE:** Negative.

**ERIC LANDER:** What happens if I do it backwards? Bad. Everybody does it backwards in the lab at some point. So then you do this. The DNA goes this way and it turns out that agarose is this matrix of cross links and all that. And DNA molecules wiggle through like a snake, through this matrix, and little ones go much faster than big ones.

So what happens is, little DNA molecules move faster than big DNA molecules. And if I stop at some point, and I add a dye that sticks to DNA, and look at it with a fluorescent light, I can actually see where the DNA molecules are. So I've grown up a ton of this plasmid, lots of it, couple of micrograms of it. I cut it, I add it to the top of my well here, I turn on the electricity, and it wiggle, wiggle, wiggles through.

This is the vector, and this is the insert. It might be actually, that I found another plasmid that could do this too, that rescued yeast. It had a vector and maybe at a different insert. Maybe I had one, I don't know why, that had the same insert right there. In any case, the first characterization that I would do is look at the size of my insert.

How big is that insert? I might find out that the insert is one kilobase, 1,000 bases, or one KB, or 2 KB or something like that. That's a first order of characterization. Now, of course, what I really want to know about the insert, sequence. I want the DNA sequencing insert, right? Let's not just mess around with, it's got 1,000 bases, I want to know what are those bases in what order. So that means I have to invent DNA sequencing.

All right. So how am I going to invent DNA sequencing? This is a little tough. I have my piece of DNA. I cut it out. Here's my piece of DNA, 5 prime to 3 prime, 5 prime to

3 prime. Let's give it a sequence ATTAAGAATGCAT, et cetera. DNA sequencing is actually remarkably cute. I take a primer just like Kornberg did back when he was discovering polymerase.

Here's my primer, let's say. Here's my primer. Here's the matching template. And what did Kornberg teaches us? He taught us that if we add polymerase, what will polymerase do? It'll add the right bases, right? TACGTA onward.

If we could look really fast, maybe we could just see it going by. But we can't look really fast. If we just add polymerase, it'll just zip up and add all the bases.

But someone had a clever trick. The clever trick is this. Suppose I threw in a smidgen of defective Ts, Ts that couldn't be extended for some reason? I'll call that T star. What happens if a defective T goes in there? Can't be extended anymore. And what happens then to my reaction? It comes to a halt.

Ah, but what happened if I put in a good T? Then instead, what would the next base be? A. Then what would happen? C, G. Then what's next? T. And what would happen if I put in a defective T? It would stop.

Now, maybe instead the defective T didn't go in. And it would go on, and the next T that it encounters, maybe it would put a defective T there. Or maybe it would put a defective T there. Now, of course, I just have a little smidgen of defective T, right? So which of these will happen? Will it sometimes stop at the first T? Sometimes at the second? Sometimes at the third?

I'll actually get a series of molecules. What will the lengths of those molecules tell me? Where the Ts are. Kind of cute. This is very cute. So if I do that reaction, I'll call that the defective Ts, I could see that there is some molecules, stop here at this very small size, some stop here, some stop here, some stop there, some stop there. And what have I just learned? The sizes, the lengths of the fragments that end in T.

But that's not good. I also want to know the As.

**AUDIENCE:** Do it with defective As.

**ERIC LANDER:** Do it will defective As, right? I'll just use defective As. And then I'll see the As stop there. And then I'll do it with defective Cs, and I'll do it with defective Gs. And I'll see exactly where the molecules are, the lengths of the molecules that end in A, T, C, and G.

And I've just read the sequence, haven't I? Pretty cool. This actually is so cool, it won a Nobel Prize. I mean cool usually means won a Nobel Prize or something like that in this class. So actually, how do I visualize the little letters there? So the little fragments on the gel? I visualize the little fragments back in ancient days, yes?

**AUDIENCE:** How do you know [INAUDIBLE] necessarily be inserted in every place where there is a letter?

**ERIC LANDER:** I'd make the right ratio. I'd choose a ratio of defective to good Ts. Like if I choose 1% defective T's, then about 1% of the time, I put in the defective T it turns out. It turns out to work OK. I just put in a smidgen of defective Ts, like 1%, and 1% of the time it stops here, and 1% here, and 1% there.

So and then, of course, to see them in old days, what people did was they made these primers radioactive, and these radioactive primers here within the gel would be held up against a piece of x-ray film, and you would see where little radioactive bands were. That's how you could visualize the primers, because it turns out you have too little primer to see it even with my fluorescent dye.

But a little radioactivity, you add radioactivity, you take the gel off, you put on a piece of, I guess you've got to wrap it with Saran. You'd better be standing behind a shield. So you're standing behind a plastic shield, you reach around the shield, you wrap it with Saran Wrap. You then put x-ray film on.

You are you doing this in a dark room. You then seal it up. You put it in the freezer for two weeks. You take it out. You develop it. You get your magic marker, and you follow where all the little bands are. And that's DNA sequencing.

It's cool. It won a Nobel Prize. It's also like really tedious. So little tricks were done to speed it up, some nice pieces of engineering. Instead of visualizing our DNA, faster DNA sequencing, what happens instead?

What people did, which was very cool, was they simply took the As, Ts, Cs and Gs, the defective ones, they attached a colored dye to each one. The As were green, and the Ts were blue and the Gs were red. Now, when I run it out, I can do it with fluorescence. And I can see the Ts all glow one color, and the As glow another color.

Now, it turns out, of course, if they're glowing four different colors, I don't actually need to run them in separate lanes. I could run them in the same lane, right? So I could run them all in a single lane, and I'd see Ts and As, and Gs.

And now, instead of putting this up against the piece of x-ray film, I could do this in a very thin tube called a capillary. And instead of holding it up to an x-ray film, I could just put a laser detector.

I can turn on the electricity and watch the fragments go by and just look at it with my laser detector. Shines a light. I say, oh yeah, green, red, blue. That's the sequence. This made it a lot easier than that. I actually got involved in molecular biology when it was still that. That is a lot better.

By the way, I didn't tell you what the defective Ts were, did I?

**AUDIENCE:** [INAUDIBLE].

**ERIC LANDER:** Sorry?

**AUDIENCE:** [INAUDIBLE].

**ERIC LANDER:** Well, they gotta have something that prevents them from being extended, right? What prevents them from being extended? Do we remember how DNA is extended? We have a sugar phosphate backbone, right? Phosphate, base, 5 prime.

Where do we extend, what base? 3 prime. And what do we use to extend on? The 3

prime hydroxyl. Remember, DNA has no hydroxyl at its 2 prime position. Its normal DNA is 2 prime deoxy.

How do you think we prevent it from being extended?

**AUDIENCE:** 2 prime [INAUDIBLE].

**ERIC LANDER:** 2 prime, 3 prime dideoxy. Defective just means 2 prime, 3 prime dideoxy. See, we tell you about these structure not just so you memorize structures, but because those are the tricks we use in the trade.

Now, you get yourself some 2 prime, 3 prime dideoxies. You put in a smidgen, maybe 1%. You run your reaction. Those dideoxies have dyes, spelled differently, D-Y-E color dyes attached to them. And you sit there with your laser detector.

Turns out to be even better. If you have one capillary, you could have two. If you have two, you could have three, and the machines that were built around 1999 had 96 capillaries. I'll call it 100.

You could sit there and read 700 bases every hour. You can do this maybe 20 hours in the day because it's automated. You could do this 360 days a year. And at MIT, we had 100 machines during the Human Genome Project. And if you do the arithmetic, we could read about 60 billion bases a year, 60 billion bases a year during the height of the Human Genome Project.

Now, this is pretty cool, but it ain't nothing compared to what we're doing 10 years later. Currently, we are doing 60 billion bases every 10 minutes at MIT. There's a little trick that I'll tell you about next time of how DNA sequencing went from that, where you could do about 250 letters a day, when I first got involved in this, to that, where at a large facility like MIT, you could do 60 billion letters a day, to today when you can do about 60 billion letters in 10 minutes.

We'll touch on that next time.