Interview with Dr. Mark Adams
Daniel Lewin, The Social and Political History of Genomics, Duke University
2006

MARK ADAMS: Hi, this is Mark.

DANIEL LEWIN: Hi Dr. Adams. Do you prefer Mark or Dr. Adams?

MA: Mark is fine.

DL: Mark—okay. Um… so… yeah, I guess you sound pretty ready for the interview so--

MA: Yeah fine.

DL: Thank you for talking to me, again.

MA: Sure!

DL: So, at Dr. Cook-Deegan's orders—

MA: *Laughter*

DL: I need an--

MA: You're taking a class? Or you're in a thesis project? What's the---

DL: I'm actually a freshman here, so it's- he teaches a class for freshman, and he's actually our faculty-in-residence, which is really nice; we have access to a lot of good stuff. And his class is basically just a history of the Human Genome Project.

MA: So it's ancient history now.

DL: *Laughter*. I guess it's not quite over yet

MA: Depends on—figuring out when it started is easier than figuring out when it's over.

DL: That's true. And there's a lot to figure out from it, as I'm learning. *(Pause)* Alright, so this is Informed Consent Statement for Oral History Interviews.

The information I am about to give you and your response will now be recorded.

My name is Danny Lewin and I am a student at Duke University. I am in a course on the history of genomics that includes oral history. One goal is to produce a written transcript of interviews with important figures in genomics—like you. Some of the interviews may be archived or made public through a website.

I selected you as the person I would like to interview. The interview should last about 45 minutes. Your participation in this interview is strictly voluntary, and you may withdraw at any time. You do not have to answer every question asked. The information that you provide will be "on the record" and may be attributed to you.

This interview is being recorded and I will take written notes during the interview. The interviews that are posted publicly- which may or may not be this one- will be archived as a history resource. (I'm reading this off a sheet). If you prefer that the interview be used only for the course and not made public, please indicate this now.

MA: It's alright if this is public.

DL: Alright, great! One risk of this study is that you may disclose information that later could be requested for legal proceedings. Or you may say something that embarrasses you or offends someone else when they read it on a public website. The benefit of participating in this study is ensuring that your side of the story is properly portrayed in the history of genomics.

Dr. Adams- Mark- do you agree to the interview?

MA: Yes.

DL: Excellent! So, that's out of the way. Let's jump right into it. I think a lot of what I want to focus on in talking to you is—it's public a lot of the great things you've done, the things you've achieved, but I want to know how you get from point A to point B, how you get to the places where you can make the publications and the discoveries that you've made, so… You got your Ph.D. at the University of Michigan.

MA: Mm-hmm.

DL: And so after that you- what gets you to NINDS. What incentives do they offer to get you there?

MA: Right. So when I was finishing up-- when I was working on my Ph.D. I got interested in computers. I taught myself programming- I had actually taken some programming classes in college. But I kept teaching myself programming, and actually had a little of software company that did reference manager kinds of things, like EndNote. I was in graduate school, and was kind of interested in that of interface between biology and computing, before the web. And when I was looking for a postdoc opportunity, I saw the ad for Craig's lab in the back of Science, and he was looking for both computer programmers and postocs in the same ad, and I thought, "Well, that sounds like it would be an interesting place to work." So I sent in my information and applied for a postdoc job, and it turns out that the person who received that had been one of the very small numbers of purchasers of my reference software.

DL: Wow!

MA: And despite that, they asked me to come for an interview. *Laughter*

DL: *Laughter*. Despite that. *Because* of that.

MA: And I had a great time visiting there, Craig is obviously a dynamic person, and had some really great, great ideas, and looked like they were going to be very interesting projects. He had assembled a very interesting group of people that seemed like it would be fun to work with, and it just worked out very well.

DL: So by the time you got there, is he working on sequencing?

MA: Yes.

DL: So you come into that--

MA: So in fact, the reason that he needs the postoc postdoc, this was the time that he had applied to NIH for intramural funding to do X-chromosome sequencing from the genome office, which was run by Watson at the time. And when I accepted the postoc- he offered me the postdoc the day I was there for the interview; I don't know if he called my references ahead of time or if he just doesn't bother with that when he wants to hire somebody- he offered me the job right away, with the idea that I was going to be working on this human genome sequencing project. So I show up several months later, having finished my Ph.D. and everything, and the first thing he says is, "Oh, we didn't get that grant." *Laughter*. And I didn't know for anything what that meant, I thought, "Oh, do I still have a job?"

DL: *Laughter*

MA: *Laughter*. And he said, "Yes," fortunately. But everybody else in the lab was working on one of two projects. One was chromosome—sequencing cosmids, either from the Huntington's Disease region on 4, or the myotonic dystrophy region on Chromosome 19. And, that was grueling work at the time, I think Craig's lab was one of the best- who had automatic sequencing working the best of any labs. There might have been labs that were doing more sequencing, but I think his was probably the farthest ahead in getting the ABI sequencers to actually work. And it was thousands of reactions, and the capacity was less than 100 a day, I mean, everybody thought that was great to be able to sequence 100, get 100 sequences every day.

DL: yeah.

MA: Once all the data started coming together, and- this was- it's hard to imagine, but the data was collected onto floppy disks- 5 1/4" floppy disks

DL: *Laughter*

MA: And to assemble it meant someone sitting down in front of a computer with this stack of floppy disks-

DL: Just loading them, wow

MA: and putting them in one at a time to pull up the right reads. It was before real sequence assembly software came along. And so I'd only been there a couple of weeks working on this little technical program when Craig came back from a long trip to Japan and he'd been working on trying to find the genes in these cosmids. You know, "Let's just start sequencing cDNAs and maybe we'll find the genes that belong to this cosmid," which was of course a crazy idea because one cosmid is like .001% of the genome, but the idea that information density in cDNAs is vastly greater than the information density in the genome was right on target. So that's what I started working on, that's how EST project was born.

DL: So EST was kind of Craig's brain child?

MA: Yes.

DL: So, what--

MA: In fact, the rest of us came to it kicking and screaming, at least initially.

DL: *Laughter*. So he just kinda tells you, "I'm thinking about working with cDNAs, and then, small parts of those." Do you—so, you say kicking and screaming, but you ended up working on it, and--

MA: Yeah, well, and it didn't take long it all. I was working on a double-strand plasmid DNA prep to facilitate sequencing. And cDNA clones are double-stranded. And, so it was a natural thing to try and we had several cDNA libraries in the lab that were being screened for using exon sequence from the cosmids as a probe to try to pull out the whole cDNA to get the genes annotated, and we just picked a couple of libraries- one was a hippocampus- and just started prepping clones. And within the first hundred sequences that came off, the first experiment essentially, there was the first mammalian ortholog of the notch gene in *Drosophila*, which is- there's now a small family of them- but, critically important in differentiation, in cell-cell signaling and differentiation of neuronal cell types. And, a gene called- the mammalian ortholog, human ortholog of a gene called enhancer of split, which interacts with notch in the same signaling pathway. And neither of these had been found in mammals before, and yet were critically important in *Drosophila* development

DL: Oh, so this has already been seen in *Drosophila*.

MA: Yeah, so the genes had been cloned and the function worked out in *Drosophila*, but they'd never been seen in mammals before.

DL: Wow

MA: And so, from then on, it was, "Oh yeah, we have to do this. We're going to find phenomenally interesting things, despite sequencing random cDNAs."

DL: So, the gene has the same function in mammals and *Drosophila*?

MA: Very, very similar, yes.

DL: So, yeah, I'll bet that leads to a lot more research into it. So you work on the Expressed Sequence Tags. What went into the decision to try to patent them, or apply for patents?

MA: Um, so that's a long--

DL: Long story, I'm sure

MA: Probably many things that went into that, and I'm sure that it evolved over time. So, Craig had previously filed a patent application on an octopamine receptor, which is another receptor- it's a seven-transmembrane receptor family from *Drosophila*. So the family of genes- proteins- Craig worked on before getting into genomics were these seven-transmembrane receptors, to which approximately half the pharmaceuticals on the market used them as their target. And so he was well aware of the implications of gene discovery and the importance of patenting in order to protect inventions and permit commercial development. And when we went to publish all of these sequences, we went to the patent office at NIH to ask whether any of them ought to be patented, thinking that, "oh, maybe this notch might be interesting," for example, even though we only had partial sequence. And Craig and I went to this first meeting with Reid Adler, who had just taken over as the head of the Office of Technology Transfer, (I'm not sure what its official name is or was) at NIH, who had been studying this area with a variety of biotech companies. He had been thinking about not necessarily just cDNA patenting. Randy Scott of Incyte read our *Science* paper and THEN shifted Incyte's focus to EST sequencing. I think several groups really started thinking about cDNAs at about the same time, but it was really our group- Craig's group- that shows that, yeah, it was really going to be a good idea to do it at ridiculous scales, ridiculous for the time, and then continuing on to ridiculous. So, it was really Reid Adler's decision to try to file patent protection on all these things. And I think at that time we really didn't appreciate the level of- I certainly didn't- the level of opposition that there would be to that decision. And there's another thing we didn't appreciate the level of opposition to which was the EST concept in general.

DL: Right.

MA: So all papers that are submitted for publication from the NIH have to be submitted to the internal review. And the internal reviewer who read our Science- 1991 Science paper- recommended that it not be published, and the data not be deposited in GenBank.

DL: At all.

MA: At all. Because it would contaminate GenBank. It was incomplete and not perfectly accurate sequence, it wasn't a final product. And this reviewer thought that it was inappropriate and not science. Not good science. So that was- I was frankly much more concerned about that issue, about not getting my first author science paper less than a year after starting my postdoc than I was with patenting. I didn't fully appreciate the ramifications, and it seemed like it was a long way off and somebody else's problem.

DL: So this seems to be a continuing controversy about the methodologies of you and Craig, and other scientists. So what went into debates about the quality of science that goes into making ESTs?

MA: I think that that individual who gave us the negative review was really in the minority.

DL: In the minority?

MA: An increasingly vanishingly small minority. The value of ESTs I think were immediately grasped by most everybody. And it was at a time- this is a continued *(RECORDING SKIPS)* when I think people would rather have partial, incomplete data as a stepping-stone to what they want to accomplish than no data at all.

DL: Especially when you're using floppy disks to load them into--

MA: Yeah. Right. To be able to do a GenBank search and be able to find your gene of interest or a bit of your gene of interest- people at the time were using degenerate PCR to try to clone these orthologous genes, and it just didn't work very well, and it was expensive and time consuming, and if you can get an exact probe to use to go in and get the full thing, that was even better. That was by far way better.

DL: So, you're at NINDS, and you and Craig and a bunch of other scientists end up forming the Institute for Genomic Research. So what went into that, leaving the NIH and going to- see, I pronounce it "tiger", I haven't heard it, actually-

MA: "Tiger" is right.

DL: TIGR? Okay. So, yes, what goes into you all leaving and creating TIGR?

MA: Right. So there's the leaving and the starting part, and obviously they go together, but there was both push and pull, I think. So by '92, things had gotten, we had really- the ESTs were really quite successful. The sequencing was working very well, we were able to do more and more of it. And we started to want to characterize sequences from many different human tissues. However, like for instance cancer is an obvious one. The Cancer Institute did not believe it was in the purview of the Neurology institute to study cancer,

unless it was like brain cancer. So, we were not- we were actively discouraged from sequencing colon cancer ESTs, for example. We were discouraged from sequencing heart ESTs, by the heart institute. And so it seemed increasingly clear that--

DL: Under the guise of--

MA: The NIH sort of mold didn't fit with what we were trying to accomplish.

DL: So is this a mutual agreement of the scientists who are trying to work on the ESTs?

MA: Really, Craig- these were discussions at Craig's level dealing with NIH and trying to get an increase in funding to support more EST sequencing. And the message coming back "Sequence all the ESTs you want as long as they're from brains." And I think that was very frustrating to him- he doesn't like being constrained.

DL: Right

MA: *Chuckles*. And that was about the time that- it was the beginning of the biotech boom. I really think that ESTs contributed to that in large part with Incyte making an enormous investment in that. And then HealthCare Ventures came to Craig and said, "We want to talk to you about setting up a company to sequence ESTs." And, Craig- again brilliant, said, "I don't want to start a company, but if you'll fund me to work at a non-profit research institute, you can have the intellectual property." He didn't want it. He didn't want to have to make money out of this, he wanted to do the science.

DL: Just straight science.

MA: Yep.

DL: And so, you hear this, you just want to continue sequencing ESTs or is there another reason for wanting to leave?

MA: I think that was really the main reason, we wanted to be able to scale up and do this in a way that it was clear was not going to be possible at NIH. For a whole variety of reasons: we had no space, it was hard to hire people, it was hard to change funding levels, let alone the programmatic issues.

DL: So then at TIGR, to my understanding you were the director of DNA sequencing?

MA: Yes.

DL: Is that- does that take on different roles than you had at NINDS.

MA: Yes. At NINDS I did get- well, I got promoted to like a staff position. But, my role with respect to sequencing was a little ambiguous. And Craig always kinda liked that, he liked flat management structure, and he liked everybody to feel like they were in charge,

in the sense that they were obligated to make everything work. So he didn't want to say, "Okay, you're the person and this is going to succeed or fail on your back." He was much more teamwork-oriented than that. Sort of a shared responsibility. And so it was a little unclear, I didn't supervise anyone at NINDS. Maybe at the very end I supervised more formally the sequencing group. But there was a very senior technician, and she really ran the place, Jenny Kelley. And so when we came to TIGR it was a little more formal- that I needed to build out this group and equip it and get a real staff that was going to be quite a bit larger than what we had at NIH. We were talking about- I guess we had 7 sequencers when we left NIH and we were going to have 100. So, pretty big scale-up. So, yeah, I had more formal responsibility for that. And we brought in many more technicians.

DL: How did you recruit the technicians?

MA: Boy, I remember much more clearly when we started Celera than when we started TIGR. But, we had ads in the Washington Post, and it was pretty easy. We got a number of good people, we recruited a lot of recent grads, rather than people who had a lot of experience. But we had several who had worked as medical technologists previously, like doing blood testing and clotting assays and things like that. And they were- that's sort of a mentality of person who's willing to do routine work very carefully and enjoyed that. So it's kind of a mixture of people who are more research-oriented and people who are more clinical lab-oriented.

DL: And so, you personally, are you still doing some of the sequencing on your own?

MA: So one of the dirty little secrets here is I never really did a sequencing reaction myself until I came to Case.

DL: *Laughter*

MA: *Laughter*. Yeah, I never ran the sequencers.

DL: So, what did you do in terms of—I guess I'm not very technical on this, but--

MA: Well, we had excellent technical folks who ran the instruments. They did a lot—I worked with them on organizing the process flow and worked with the architects and engineers on designing the space and working out the protocols and defining how to do the quality control, assess the performance, find areas where we could improve the performance and lower the costs, all of these kinds of things I worked with the sequence group to do that. And then I also worked with the software group to develop the methods for doing the data analysis. And these sort of dual roles were what I did all the way through TIGR and Celera. More direct responsibility- managerial responsibility- for the sequencing group, and generally speaking, less responsibility for the software people, although a lot of influence.

DL: So you're kind of directing the technicians as to what (*pause)*- I'm sorry, I guess I'm not really fully grasping. Is it an administrative role, or are you directing the science, or

both and I'm a little confused?

MA: I think both with a little less emphasis on administrative. I did participate in the job interviews and trying to choose the right people, but a lot of the day-to-day operations were under Jenny Kelley.

DL: Still under TIGR they were?

MA: Yeah. And then she left TIGR before Celera got started.

DL: Okay.

MA: But sort of the operation side of things, I've always had really good people who dealt with a lot of the day-to-day stuff. So I worked much more on- I wouldn't necessarily call it big picture, because there were a lot of nuts and bolts involved- but how do we make this work better? At every step of the way. And that's a constant- yeah, I don't think there was ever a time- there was rarely a time where things were like, "okay, we're set. We're not going to change anything now." That's part of the history of the genome project is constant evolution of the technology.

DL: At first you can only find so many genes a day, and all of a sudden, it's exponentially increasing.

MA: Right, there's always pressure to do more with less money, and both from the less money and the more standpoint. I mean, we always want to—it became a race to collect the intellectual property or find the interesting target, to get the next data set ready for publication. And always to save money. And that's still true. There's the $100,000 genome, the $10,000 genome, the $1,000 genome projects that are still actively being pursued.

DL: Mm-hmm. So, you work at TIGR, and then what goes into the process—eventually people end up at Celera. Is that basically an extension of TIGR? Is it a similar story of going from NINDS to TIGR?

MA: (*Pause*)

DL: I guess, kind of a convoluted question. How do you get from TIGR to Celera, is what I'm trying to ask?

MA: Right. So, there's a long history of TIGR that's interesting in and of itself. And I don't know if you've read either the Jamie Shreeve book, or the Ingrid Wickelgren book on the genome project. Both of those have a good history of that- the relationship with HGS forming and then being dissolved. Working with Ham Smith to sequence the *Haemophilus* genome, which really was the beginning of setting off towards doing human genome sequencing.

DL: I'm sorry, which genome was that?

MA: *Haemophilus influenzae*

DL: Influenzae, okay.

MA: The first genome of a free-living organism. Sequenced in 1995. So we were- TIGR was focusing on- I had a human genome grant at TIGR, part of the- one of the pilot groups that were scaling up human genome sequencing and there was still a lot of microbial sequencing going on and Craig all along had a very good relationship with Applied Biosystems. We paid a lot of attention to quality and tried very hard to get the most out of the sequencers, and I think among the sequencing centers—they all had to work with ABI—but I think our relationship with ABI was a little more productive than some of the other sequencing centers.

DL: Was it something in particular?

MA: Well, a lot—everybody was frustrated with ABI as a closed system, the prices, the support, you know, a variety of things. And we, for whatever reason, had good relationships there and tried very hard to work things out together. That's not to say that other labs didn't, but we always—Craig always had a good relationship with the people at ABI. And so when they had a new technology coming along, they called us up. And this was coming at a time—again, you know, the biotech revolution was booming in the late 1990's, and someone there, whether it was Tony White or Noubar Afeyan or Mike Hunkapiller or somebody else, I don't know, had the idea: "Well, if we have this technology, why should we just sell it to customers, why can't we use it to gain some extra advantage? Extra value, as the corporate term goes. Let's start a company and sequence the human genome with this new great technology that we have developed." And so they called Craig and talked to Craig and me and I just laughed. "You want to start a company to sequence the human genome?" It was just incomprehensible to me, not having seen the technology. And we were struggling over three years, my group, which was—you know, one of the six groups that had a shot at being in the running to sequence the genome. We had new pilot project grants. We were going to sequence 11 megabases in three years. .3% of the genome. And the idea of starting a company to do the whole thing was just ludicrous. But, they kept calling, and so Craig and I went out to visit in February of '98, and saw the technology, what was going to be the 3700.

DL: So these are new sequencing machines?

MA: The new capillary—96 capillary sequencers

DL: Okay.

MA: And they also showed us a bunch of other technology. So the sequencing is just one part of it, okay, but you have to feed the sequencer with Sanger reaction products that are fluorescently labeled. Those have to come from templates. Those templates have to come

from libraries. And those libraries have to come from somewhere. And they showed us technology that could be—potentially address each one of those up front steps to automate the millions of sequencing reactions that were going to be necessary to sequence the human genome. To feed the 3700. So just having the 3700 was not adequate to sequence the genome. But they had all these other things that looked really interesting too. So, we did some back-of-the-envelope calculations and said, "Yeah, we could do that. And it would only take 200 of them." *Laughter*

DL: *Laughter*. Wow. So, 200 to sequence the whole human genome in 2 years, was the--

MA: Yes. Yep. That was the idea.

DL: Wow.

MA: And that—so they went very, very rapidly from there. They went back to their board and got approval to fund it. They were most definitely interested in investing some of their capital in trying to move up the value chain, I think. And it was—when we left that meeting in February '98, we did not know what form it was going to take, we just knew we wanted to do it. I thought that it might be possible to do it in the context of TIGR, but it turned out for essentially tax—I guess business reasons that it was not possible to just fund TIGR to do it in whatever way shape or form. It would have jeopardized TIGR's tax-exempt status. And there were other issues as well. So it turned out it was going to be a company, and it took another long while to figure out that it was going to be a tracking stock rather than Applied Biosystems itself and the train had way left the station by the time all of that was figured out. That's my recollection of it.

DL: So you're just kind of thrown into this company, but there's not--

MA: Yep.

DL: so much planning behind it.

MA: Yep. *Laughter.*

DL: *Laughter*.

MA: And it was for the—you know, we went for the science. We went because we wanted to sequence the genome. And it was quite clear that it was going to happen.

DL: You worked out the rest of it after you'd gotten there.

MA: Right.

DL: And you had said earlier that Craig didn't like so much to be constrained, but now he's kind of—he has to watch out for a bottom line for ABI and the other investors in Celera. So how did that conflict work out?

MA: I think Craig was Celera's best salesperson, because he could articulate more clearly than any of the businesspeople what the value of having the genome information was. And, particularly early on, for the early access customers who had seen his track record, they believed it. And I think it was right, that we offered the promise of accelerating their access to genomic information. I should say just as part of finishing this transition that the 3700 did eventually work, but none of the other technology that ABI showed us that would facilitate automating the whole process of sequencing the genome did it ever work. And so we spent the first six months at Celera basically building an operation to do sequencing at 50 times the throughput that was currently operating at TIGR at the time that we left. 50 times higher throughput. And it involved changing every aspect of all of the up-front sample processing. And a very small R&D group of 3 or 4 people did Herculean work to do all of that and work out all those protocols.

DL: Is this including you?

MA: I didn't work in the lab. But I'll take a little credit for helping point them in the right direction.

DL: Alright.

MA: Yu-Hui Rogers, who is now head of the Venter Institute's sequencing facility was really a key leader there.

DL: So you all—except for the 3700s, you're saying, everything else you had to recreate.

MA: Yep.

DL: So, was there the capacity to do this all along and you weren't looking to map the whole human genome right away at TIGR.

MA: So this was a-- sequencing is a multi-step process. And the trick to scaling up the throughput is twofold: one is to recognize where the rate-limiting step is. What is it that makes it so you can't do more? And for a long time that was the number of sequence reads that could be—the lane. How many times can you load the gel and run that? Because the machines were expensive and you couldn't just buy more of them. So the whole process was developed in order to feed a certain number of machines. And so then they had 24-lane and then 48-lane and then 96-lane. So there was a nominal increase in throughput that was at a somewhat measured pace. But here we were told you can have essentially an infinite number of machines. The lanes aren't going to be limiting. So then everything else became limiting all at the same time. But until then, there really was never any incentive to try to work out how to do 200,000 template preps a day. Because--

DL: Because you couldn't run them--

MA: you wouldn't have had anything to do with them.

DL: Okay, alright. So then your role at Celera, was that similar to what you were doing at TIGR beforehand?

MA: Yes, with a much larger role, I think, in the informal role of helping with the software development, because there was so much more of it to do. So the one group I was formally responsible for was the annotation team. So that's all very well and good, except you need something interesting to annotate. And Gene Meyers's group was responsible for assembly. And then someone else was responsible for a group that was involved in data processing- getting data from the sequencer to Gene Meyers's group. And I worked with all of them to make sure everybody was talking to everybody. I mean, the company grew so fast. I think about 15 people left TIGR to start Celera, and we had over 200 at the end of the first year. And mostly the software people did not come with a biology background, and just hadn't necessarily worked together—hadn't necessarily worked with this kind of data and these kinds of problems. So there was a lot of talking and educating to do, and I did a fair amount of that, trying to get people on the same page and help the software engineers understand the "gotcha"s that came with biological data with all of its noise. A lot of them were very unused to that.

DL: So you're in a sense teaching the software people what they're looking at?

MA: Yes. So they knew how to write an algorithm that would do any ABCD things fast, accurately. But they didn't know what that algorithm needed to do. So sequence data is of variable quality, you recognize that. What do you do with it when you find it? These sorts of things. They couldn't—you can't just tell them to write a sequence trimmer, because they don't have any concept of what that means. And Gene Meyers's group did a phenomenal job putting together the sequence assembly software, which critically depended on having data of a defined quality, and he was very worried about that all the way along. And so he worked with the folks in the lab, so that he could understand what the quality of the data was that was coming off. We both worked with the software engineers to help define that level of quality so that it wouldn't be a garbage in garbage out process.

DL: So to get, I guess, kind of technical, what defines quality data versus not-so-quality data?

MA: So there's only one way that sequence can be good, but there are many, many ways that it can be bad. *Laughter*. So there was—there is a way of assigning an error probability to every individual base in a piece of raw sequence data. But there can be either individual bases that are of poor quality, and that can be either random or it can be systematic, and we were very concerned about systematic errors because a lot of the technology was new and we didn't know in what ways it was going to fail. Then there are a whole variety of different sequence-based reasons why there might be poor-quality sequence data, like long runs of the same nucleotides one after the other. And then there were more structural ways in which the data could be bad. Mate pairing was an essential feature of the shotgun sequencing strategy, so it takes advantage of two adjacent reads in

the genome that we know about how far apart they are. But if a clone has a mixture of DNA from two different bits of the genome, then that's going to throw up—throw that whole thing out the window. So it had to—we had to check that it was a very low percentage of those double-insert clones. The whole variety of things like that, each of which needs its own way of looking for it, and then we got surprised by some things that we didn't know were going to be a problem as well.

DL: So, but you all just kept looking for the problems, found them-

MA: Right.

DL: Tried to solve them-

MA: Right.

DL: So, all the while at Celera, is there—was there a general feeling among the people working there that this was a competition, this was a race to find the whole human genome, among you all and the, I guess, well, a lot of other groups.

MA: I would say yes and no. I think, at least, I certainly, the other scientists recognized that it's all the same genome. It's not like one group is going to "get the genome" and the other is not, assuming that we're both successful. I think the real—the race was on behalf of our customers to get them something that was usable as rapidly as possible, because they were paying for it. And the PR race, it became a PR race. And neither of us—neither we nor the publicly funded group wanted to appear to lose that race. And neither of us could afford to appear that we lost that race.

DL: So, is this working kind of independently of the science or--?

MA: Largely. I would say yes.

DL: Okay. People on their own… Does it—I guess, does it, trying to please customers, does it change things if a publicly funded project makes it publicly available—the sequence that they found?

MA: Sure. Partly what our customers were paying for was access to data that wouldn't be available anywhere else. And that comes in a variety of forms. It's raw base pairs, like ESTs that can be used as the beginning of more projects, certain kinds of projects like looking for seven-transmembrane receptors. And I think, I expect that all of our customers were looking for those and found them. One of the nice things about those is that a lot of them are single-exon, so once you get a little bit, you get most of it, unlike other genes, which are multi-exon and span large bits of the genome. And you get—in a genomic sequence, one read matching one exon might give you only a tiny fraction of the full gene. (*Pause*) I can't remember where I was going with that.. Oh, the value of—the levels of information. But it became increasingly clear that having assembled information was going to be more and more useful. And I think that Celera's assembly data, although

it never came along as fast as anybody wanted it to, still was better for a long time than the public assembly.

DL: So, customers were kept happily.

MA: Variably. *Chuckles*.

DL: *Chuckles*. Who was—I guess, who was—was it pharmaceuticals you were?

MA: So, Celera did initially sell to large pharmaceutical companies. They paid a lot of money and they were demanding, and I think justifiably so for the data. They could afford and assigned relationship managers to work with Celera and they made it so that the data that they were getting was distributed in their organizations to the people who wanted to get access to it. As time went on, Celera also started marketing directly to university labs and smaller biotech companies, and certainly we all—we wanted the data to get out as widely as possible, but it was the large pharmaceutical companies that, I think, their contracts that were negotiated drove a lot of the development of the product.

DL: Alright. And so, I guess, you have clients who are looking for this human genome sequence. And then, how does that work out with other sequencing projects that are going on within Celera. And, to my understanding you had a big role in the *Drosophila* sequencing project.

MA: Right.

DL: So how—so first, how did Celera balance time between human sequencing and--

MA: Right. So, *Drosophila* is 5% of the size of the human genome. At the time that we got fully up to speed, we could have sequenced *Drosophila* again in about 10 days.

DL: Ah.

MA: And, that was the argument that I kept making, Craig kept making, everybody else kept making during the 6 months that we spent sequencing *Drosophila* while we were working out all the technology bugs. While all of our customers—actually, that's not true—some of the customers were very interested in getting the *Drosophila* data. But, for the most part, they wanted the human data. And what we kept telling them was, we need to know if this works. And the best way to do that--

DL: Test it. Mm-hmm.

MA: is to sequence *Drosophila* and get it done with. Whereas-- and what's going to happen is, you get the human genome ten days later. And, if we don't do this, though, and there are problems, who knows what could happen. And so we stuck to our guns, and we sequenced all of *Drosophila* before we started doing any human genome sequencing.

DL: So then--

MA: It was the same day the last *Drosophila* plates came off the sequencers, the first human plates started going on the sequencers. And there wasn't a break or anything. And we just—it was—that was still very much in the middle of the ramp-up, in terms of capacity, that Celera was going to ultimately get to. But there were a lot of folks who were not in favor of us doing *Drosophila*.

DL: But then, at the same time there was the scientific group that got together, and I've heard it called the Annotation Jamboree a couple of times

MA: Yeah. Mm-hmm.

DL: How did that collaboration work out?

MA: It was a lot of fun, and it worked out very well. It was, I think, there was a lot of suspicion among the *Drosophila* community about Celera, but we ultimately did the right things, and the community came together in support of that, and the *Drosophila* papers were good. And the Jamboree went well. And it came off, just barely, the data was just barely available in time. But a lot happened very, very quickly because of an awful lot of people's hard work, both at Celera and the people who came to visit for the Jamboree.

DL: What was your role in planning, or working at the Jamboree?

MA: So I worked with Gerry Rubin and Craig to invite people. I mean, Gerry invited most of them, because he knew the *Drosophila* community. And then I organized all the local arrangements for food, hotels, and snacks, and all of that, and was sort of the master of ceremonies for our twice-daily meetings where we would get together and talk about all of the numerous things that we wanted to do, how to do them, who did them, what they found, that sort of thing.

DL: So, did you get to—have to direct people as to what their--

MA: Really, most people came in with a pretty clearly defined role. And again, in the Craig management fashion, people got together and figured out what to do. And I was one of the people who helped do that.

DL: Alright. And I guess one last question is throughout most of the time that--

MA: Well, I should say that for the most part, the roles were pretty reasonably well defined. We knew who our gene annotation group was, and we knew who the software group was, and they talked to each other and if there was a problem, well you go to those two guys, you know. I helped steer, I guess.

DL: Still a structure to it.

MA: Right.

DL: And so, yeah, the last question I'm looking for is throughout most of this time you have been in close proximity with Craig Venter and I think James Shreeve mentioned that some people said you were attached at the hip. And, I was just wondering like what kind of a relationship you two developed and—you know, throughout your times working together.

MA: It was good, I mean we had a good relationship. He was obviously very busy, and didn't *(pause)* It's hard to describe, but, he was both—he wanted to be deeply involved in the details sometimes, so I wouldn't—he was not by any means a micromanager, most of the time. So he gave me a lot of latitude, but he expected me to get stuff done, too. But he also wasn't a slave driver kind of boss or anything like that. He was very inspiring. He set very high expectations, and—you know, extremely high expectations, but then he didn't come around and rap your knuckles when you didn't reach them. He wanted to know why, and he viewed that as an opportunity to change the way things worked. And so I think that's the—if I could say anything about him, that's sort of what it was—Completely unrealistic expectations, and then put around him a group of people who could fulfill them. He instilled the confidence in us that these incredibly ambitious projects absolutely could be accomplished. And we didn't make every one, but we made a lot of the important ones.

DL: Alright, well, I think that about wraps it up, and my recording device is almost out of batteries

MA: Okay.

DL: So, thanks again for talking to me.

MA: Sure. Glad to.

DL: Yeah, so, thank you.

MA: Alright. If you have any more follow up or anything, you know where to find me.

DL: Alright. Thanks again

MA: Take care.

DL: You too.

MA: Bye.