This PDF packet contains an e-mail chain, edited as a Word document and then re-converted o PDF, for ease of reading and understanding (pp. 1-4). The original email chain appears at the end of the packet (pp. 5-8).

Maynard V. Olson <mvo@u.washington.edu>
To: Kathryn Maxson <kat.maxson@duke.edu>
Cc: Bob Cook-Deegan <bob.cd@duke.edu>
Re: Fwd: Congressional record of July 1998 hrg

Kat,

Certainly, you may post this e-mail exchange in a public archive without restrictions on access. Great news about Princeton. Even I am aware that this is a top program in the history of science. Keep in touch.

Maynard

On Thu, 16 May 2013, Kathryn Maxson wrote:

Dear Maynard,
I hope you are well!
We are well along in organizing our Bermuda files for deposition into a public archive.
This email (fwd below), in which you embedded some thoughts in February 2012, would be very useful to scholars of the HGP, I think.
Do I have your permission to post it, under the heading "2012 some thoughts on HGP from Maynard Olson"?

Thanks!
Best,
Kat Maxson

PS: I'll be starting at the PhD Program in History of Science at Princeton this fall.
Thought you might be interested in knowing, since you were there for some of the frenzy leading up to my applications.

Begin forwarded message:

From: Robert Cook-Deegan <bob.cd@duke.edu>
Subject: Fwd: Congressional record of July 1998 hrg
Date: February 8, 2012 6:10:32 PM EST
To: Kathryn Maxson <kat.maxson@duke.edu>
Reply-To: bob.cd@duke.edu

Meant to cc you in the loop.
BCD

---------- Forwarded message ----------

From: Maynard Olson <mvo@u.washington.edu>

This PDF packet contains an e-mail chain, edited as a Word document and then re-converted o PDF, for ease of reading and understanding (pp. 1-4). The original email chain appears at the end of the packet (pp. 5-8).

Date: Wed, Feb 8, 2012 at 11:23 AM
Subject: Re: Congressional record of July 1998 hrg
To: Robert Cook-Deegan <bob.cd@duke.edu>

Bob,
See my *embedded comments*.
Maynard

On Tue, 7 Feb 2012, Robert Cook-Deegan wrote:
Here it is.

To follow up on Thursday night's conversation, and to clarify my thinking, one or two further thoughts occurred to me.

Given that most gaps that are hard to close are in repeat sequences, and those sequences in telomeres and centromeres may not be doing the informational work, but may be more related to 3-D structure, attachment to cellular organelles, and other non-coding functions, how "catastrophic" are they? Doe they get in the way of most uses of sequence?
Most uses? Probably not. It is in the nature of cream skimming that it captures a lot of the good stuff at low cost. However, in the human, it is a misperception that we are just talking about specialized genomic compartments. Neither Craig nor I were talking about the problems associated with highly repetitive sequences such as those at telomeres and centromeres. Neither the public project nor anyone else was targeting these sequences in a serious way, then or now. My hundred thousand gaps were distributed fairly uniformly across the genome. While it is true that they are more likely to be in non-coding than coding regions, keep in mind that >98% of the genome is non-coding. When the average gene is in multiple pieces, genome sequences lose utility for many purposes. Just one example would be every step of the GWAS process (fleshing out dbSNP, building the HAP map, interpreting population-specific effects on linkage-disequilibrium, choosing a covering set of sentinel SNPs, interpreting whatever associations are found). Sequencing the whole gene-containing portion of the Y chromosome, an exercise that yielded rich evolutionary and medical insights, pushed even the clone-by-clone methods to their limits--WGS just left a pile of debris.

Can you think of an example of a scientific project that got halfway and stopped because the "oomph went out" of the project as a subordinate but sexier goal was accomplished? I know this was a commonly voiced concern; I just don't know how real the danger is.
I will not try to generalize over the vast variety of scientific projects. In genomics, the risk was and is certainly real. Indeed, in the headlong rush into Illumina sequencing, there are many areas in which the "oomph" has drained out of efforts to understand biology whose genomic underpinnings are hard to get at with the new methods. A good example would be HLA associations, the mother of all genetic associations. Thirty years after Dausset's discovery, we still do not know whether the strong associations between type I

This PDF packet contains an e-mail chain, edited as a Word document and then re-converted o PDF, for ease of reading and understanding (pp. 1-4). The original email chain appears at the end of the packet (pp. 5-8).

diabetes and HLA haplotypes involve events that occur in the thymus or the pancreas. Most of the reason for this lacuna is that the genomics needed to solve the problem requires a kind of persistence that the system rarely musters. Happily, one time this level of persistence was achieved was in the HGP.

Finally, seems to me that the arguments that were strong at the time were basically:
A) We have a goal and we have a process that will reach it. Don't pull back now.
B) There's a new strategy. It is basically orthogonal to ours. It can build on ours but not vice versa. It may or may not work, and there is good reason to think there will be holes in the map at the end. The "not vice versa" point is the critical one.

The policy decision for a Member of Congress is not all that hard. You keep on with publicly funded effort; you figure the private effort will use private $ so is not so much your concern. That seems to be more or less what happened; indeed, Wellcome placed a bigger bet and NIH and DOE kept on going, with no appreciable budget cuts or admin interference that I'm aware of. Celera also proceeded.
You could judge better than I whether there was real political risk. There still was some residue of Gingrich I in the Beltway at the time--sell the National Parks, privatize education, rely on Big Pharma to improve health, bet on the ingenuity of the dot com's.

Some questions in my mind:
1. This is obviously wasteful to have two efforts when there could be one. But that's completely rational if there was an active technical dispute and both options get pursued. Quite common in biological and social systems; value of redundancy and overlapping strategies. Somewhat inefficient, but many advantages: urgency and competition between the two approaches, if one fails the other might work; if the benefits are big, the short-term costs are dwarfed by them.
Absolutely true, as anyone in the private sector will confirm. We do not have great disk drives because we put all our disk drive engineers under one roof responding to one set of incentives. In the Soviet Union, there was a remarkable ability to pursue top-down projects judged to be in the national interest. The problem was that only a modest number of activities lend themselves to this approach. Even the Manhattan Project involved a lot of bet hedging, particularly on approaches to uranium enrichment and commitment to the radically different engineering challenges posed by uranium and plutonium bombs.

2. Was there harm from doing a "draft" sequence, aside from the added expense?
No, the "draft" was a good idea. There was not much added cost and quite a lot of utility. I think the only concerns about the "draft" related to the risks, discussed above, that it would lead to a loss of momentum. Francis's commitment to a quality sequence was always suspect.

3. Were there opportunity costs from doing a draft sequence?
No, nothing significant.

This PDF packet contains an e-mail chain, edited as a Word document and then re-converted o PDF, for ease of reading and understanding (pp. 1-4). The original email chain appears at the end of the packet (pp. 5-8).

4. How does an objective observer decide in the end whether the optimal strategy would have been public project only, public project + Celera, or no public project + Celera. In a way, the decision re Celera is not a policy decision, since it's basically PE-Applera-whateveritsnamewasthatweek's investment decision. The policy decisions are: continue public project, continue public project with draft sequence, discontinue public project. Option 3 seems highly risky and unwise given arguments A and B above. So how to discern wisdom of diverting resources to draft sequence?
A pure Celera world would have been a disaster. Eventually, the scientific community would have recovered its balance, but it is difficult to over-state the scientific and policy ramifications that would have ensued. I do think it is obvious that the Celera challenge accelerated the public effort significantly (and said so in my JMB article). John Sulston disagrees, but he did not have to deal with the US funding system.

There is another set of questions that should have a technical answer about the technical dispute. What are the advantages and disadvantages of the Celera v public domain sequence? Public project has huge advantage of being open and available. Heidi Williams's paper shows that matters for advancement of science, at least, and it's probably more or less the same story for technology and application, although would be harder to prove. A 30% advantage in access to and utilization of a data resource is a pretty big deal.
It is impossible to imagine doing science in a world where private players have long-term control of data that play central roles in major fields. We really are talking about the "evil empire" here. Science may often disappoint, but virtually all other areas of our collective lives disappoint more often and more alarmingly.

Just some thoughts. Obviously fuzzy and unfocused.

The same applies to my responses.

BCD
==========

Robert Cook-Deegan, MD
Genome Ethics, Law & Policy
Institute for Genome Sciences & Policy and
Sanford School of Public Policy
Duke University, Box 90141
304 Research Drive
Durham, NC 27708-0141
919.668.0790
gelp@duke.edu
http://www.genome.duke.edu/directory/faculty/cook-deegan/

Maynard Olson <mvo@u.washington.edu> May 26, 2013 7:28 PM
To: Kathryn Maxson <kat.maxson@duke.edu>
Cc: Bob Cook-Deegan <bob.cd@duke.edu>
Re: Fwd: Congressional record of July 1998 hrg

Kat,

Certainly, you may post this e-mail excahange in a public archive without restrictions on access. Great news about Princeton. Even I am aware that this is a top program in the history of science. Keep in touch.

Maynard

On Thu, 16 May 2013, Kathryn Maxson wrote:

> Dear Maynard,I hope you are well!
> We are well along in organizing our Bermuda files for deposition into a public archive. This email (fwd below), in which you embedded some thoughts in February 2012, would be very useful to scholars of the HGP, I think.
> Do I have your permission to post it, under the heading "2012 some thoughts on HGP from Maynard Olson"?
> Thanks!
> Best,
> Kat Maxson
> PS: I'll be starting at the PhD Program in History of Science at Princeton this fall. Thought you might be interested in knowing, since you were there for some of the frenzy leading up to my applications.
> Begin forwarded message:
>
> From: Robert Cook-Deegan <bob.cd@duke.edu>
> Subject: Fwd: Congressional record of July 1998 hrg
> Date: February 8, 2012 6:10:32 PM EST
> To: Kathryn Maxson <kat.maxson@duke.edu>
> Reply-To: bob.cd@duke.edu
> Meant to cc you in the loop.
> BCD
> ---------- Forwarded message ----------
> From: Maynard Olson <mvo@u.washington.edu>
> Date: Wed, Feb 8, 2012 at 11:23 AM
> Subject: Re: Congressional record of July 1998 hrg
> To: Robert Cook-Deegan <bob.cd@duke.edu>
> Bob,
> See my embedded comments.
> Maynard
> On Tue, 7 Feb 2012, Robert Cook-Deegan wrote:
>
> Here it is.
>
> To follow up on Thursday night's conversation, and to clarify my thinking, one or two further thoughts occurred to me.
>
> Given that most gaps that are hard to close are in repeat sequences, and

> those sequences in telomeres and centromeres may not be doing the informational work, but may be more related to 3-D structure, attachment to cellular organelles, and other non-coding functions, how "catastrophic" are they? Doe they get in the way of most uses of sequence?

Most uses? Probably not. It is in the nature of cream-skimming that it captures a lot of the good stuff at low cost. However, in the human, it is a mis-perception that we are just talking about specialized genomic compartments. Neither Craig nor I were talking about the problems associated with highly repetitive sequences such as those at telomeres and centromeres. Neither the public project nor anyone else was targeting these sequences in a serious way, then or now. My hundred thousand gaps were distributed fairly uniformly across the genome. While it is true that they are more likely to be in non-coding than coding regions, keep in mind that >98% of the genome is non-coding. When the average gene is in multiple pieces, genome sequences lose utility for many purposes. Just one example would be every step of the GWAS process (fleshing out dbSNP, building the HAP map, interpreting population-specific effects on linkage-disequilibrium, choosing a covering set of sentinel SNPs, interpreting whatever associations are found). Sequencing the whole gene-containing portion of the Y chromosome, an exercise that yielded rich evolutionary and medical insights, pushed even the clone-by-clone methods to their limits--WGS just left a pile of debris.

> Can you think of an example of a scientific project that got halfway and stopped because the "oomph went out" of the project as a subordinate but sexier goal was accomplished? I know this was a commonly voiced concern; I just don't know how real the danger is.

I will not try to generalize over the vast variety of scientific projects. In genomics, the risk was and is certainly real. Indeed, in the headlong rush into Illumina sequencing, there are many areas in which the "oomph" has drained out of efforts to understand biology whose genomic underpinnings are hard to get at with the new methods. A good example would be HLA associations, the mother of all genetic associations. Thirty years after Dausset's discovry, we still do not know whether the strong associations between type I diabetes and HLA haplotypes involve events that occur in the thymus or the pancreas. Most of the reason for this lacuna is that the genomics needed to solve the problem requires a kind of persistence that the system rarely musters. Happily, one time this level of persistence was achieved was in the HGP.

> Finally, seems to me that the arguments that were strong at the time were basically: A) we have a goal and we have a process that will reach it. Don't pull back now.
> B) There's a new strategy. It is basically orthogonal to ours. It can build on ours but not vice versa. It may or may not work, and there is good reason to think there will be holes in the map at the end.

The "not vice versa" point is the critical one.

> The policy decision for a Member of Congress is not all that hard. You keep on with publicly funded effort; you figure the private effort will use private $ so is not so much your concern. That seems to be more or less what happened; indeed, Wellcome placed a bigger bet and NIH and DOE kept on going, with no appreciable budget cuts or admin interference that I'm aware of. Celera also proceeded.

You could judge better than I whether there was real political risk. There still was some residue of Gingrich I in the Beltway at the time--sell the National Parks, privatize education, rely on Big Pharma to improve health, bet on the ingenuity of the .com's.

> Some questions in my mind:
> 1. This is obviously wasteful to have two efforts when there could be one.

> But that's completely rational if there was an active technical dispute and both options get pursued. Quite common in biological and social systems; value of redundancy and overlapping strategies. Somewhat inefficient, but many advantages: urgency and competition between the two approaches, if one fails the other might work; if the benefits are big, the short-term costs are dwarfed by them.

Absolutely true, as anyone in the private sector will confirm. We do not have great disk drives because we put all our disk-drive engineers under one roof responding to one set of incentives. In the Soviet Union, there was a remarkable ability to pursue top-down projects judged to be in the national interest. The problem was that only a modest number of activities lend themselves to this approach. Even the Manhattan Project involved a lot of bet-hedging, particularly on approaches to uranium enrichment and commitment to the radically different engineering challenges posed by uranium and pultonium bombs.

> 2. Was there harm from doing a "draft" sequence, aside from the added expense?

No, the "draft" was a good idea. There was not much added cost and quite a lot of utility. I think the only concerns about the "draft" related to the risks, discussed above, that it would lead to a loss of momentum. Francis's commitment to a quality sequence was always suspect.

> 3. Were there opportunity costs from doing a draft sequence?

No, nothing significant.

> 4. How does an objective observer decide in the end whether the optimal strategy would have been public project only, public project + Celera, or no public project + Celera. In a way, the decision re Celera is not a policy decision, since it's basically PE-Applera-whateveritsnamewasthatweek's investment decision. The policy decisions are: continue public project, continue public project with draft sequence, discontinue public project. Option 3 seems highly risky and unwise given arguments A and B above. So how to discern wisdom of diverting resources to draft sequence?

A pure Celera world would have been a disaster. Eventually, the scientific community would have recovered its balance, but it is difficult to over-state the scientific and policy ramifications that would have ensued. I do think it is obvious that the Celera challenge accelerated the public effort significantly (and said so in my JMB article). John Sulston disagrees, but he did not have to deal with the US funding system.

> There is another set of questions that should have a technical answer about the technical dispute.
> What are the advantages and disadvantages of the Celera v public domain sequence? Public project has huge advantage of being open and available. Heidi Williams's paper shows that matters for advancement of science, at least, and it's probably more or less the same story for technology and application, although would be harder to prove. A 30% advantage in access to and utilization of a data resource is a pretty big deal.

It is impossible to imagine doing science in a world where private players have long-term control of data that play central roles in major fields. We really are talking about the "evil empire" here. Science may often disappoint, but virutally all other areas of our collective lives disappoint more often and more alarmingly.

> Just some thoughts. Obviously fuzzy and unfocused.

The same applies to my responses.

> BCD

> BCD
> ==========
> Robert Cook-Deegan, MD
> Genome Ethics, Law & Policy
> Institute for Genome Sciences & Policy and
> Sanford School of Public Policy
> Duke University, Box 90141
> 304 Research Drive
> Durham, NC 27708-0141
> 919.668.0790
> gelp@duke.edu
> http://www.genome.duke.edu/directory/faculty/cook-deegan/

--
BCD
==========
Robert Cook-Deegan, MD
Genome Ethics, Law & Policy
Institute for Genome Sciences & Policy and
Sanford School of Public Policy
Duke University, Box 90141
304 Research Drive
Durham, NC 27708-0141
919.668.0790
gelp@duke.edu
http://www.genome.duke.edu/directory/faculty/cook-deegan/