Note: Some details have been redacted for privacy.

Select e-mail correspondence in digital archive

ME **Me** XXX to Robert Waterston XXX January 15 | show details | reply all | more ▾

2 attachments: 2017 Robert Waterston, e-mail communication to authors, 17 Feb 2017.pdf (255 KB) and 2017 Robert Waterston, e-mail communication to authors 4 & 9 Jan 2017.pdf (618 KB)

Dear Bob,

Happy New Year. We have completed our manuscript on the Bermuda Principles [...]

Our correspondence early last year was invaluable to filling major gaps in the narrative. I have taken the liberty of condensing this correspondence into two PDF files, one from January and the other from February of 2017. These emails [...] provide a bulk of the documentation where we discuss the crucial period from 1994 through 1996. Rather than repeatedly citing our e-mail correspondence without any links, I wonder if you might give us permission to put these PDFs, attached to this email, into our digital archive [...]. Our readers will then be able to refer to the quotations you provided, with one stable URL per PDF. [...] I think that having a stable link to these sources cited frequently will make the references not only cleaner, but also more reliable, if you're willing.

Please let us know at your convenience.

Warmest wishes, and with enormous gratitude,

Kathryn (and Bob and Rachel)

--
Kathryn Maxson Jones
Ph.D. Candidate
Program in History of Science
Princeton University

2 attachments | save all

2017 Robert Waterston, e-mail communication to authors, 17 Feb 2017.pdf
255 KB | **preview** | open | save

2017 Robert Waterston, e-mail communication to authors 4 & 9 Jan 2017.pdf
618 KB | **preview** | open | save

RW **Robert Waterston** XXX to Me XXX January 31 | show details | reply all | more ▾

Waiting on others | x

Kathryn,

[...] Glad to hear of the progress and that our input was helpful. You certainly have my permission to put the aggregated emails into your digital archive. I read the document over quickly and didn't see any problems.

Best,

Bob

Fwd: Re: Submitted Bermuda manuscript

Note: Some details have been redacted for privacy.

RW **Robert Waterston** 2/17/17 | hide details | reply all | more ▾

from: ★ Robert Waterston
to: ★ Kathryn Jones
★ Robert Cook-Deegan
★ Rachel Ankeny
cc: ★ John Sulston
folder: /Bermuda/Final revisions
date: Fri, 17 Feb 2017 10:48:56 -0800
user-agent: reply-to:
subject: Re: Fwd: Re: Submitted Bermuda manuscript
one attachment:

Kathryn et al.,

I've attached the word doc with my suggested edits and comments. I took the liberty of sharing all this with John (cc'd) to be sure I had things straight.

In general I found the piece a good read. It brought back a flood of memories. It was also good to learn of others reactions and to hear a third party's view of the whole thing. You are too kind!

Most of my comments are focused on the parts I know best -- at the start and nearer the end. The most significant changes are on the description of the physical mapping, where you adopted Maynard's 1989 description of hybrid maps, which didn't really apply to the worm. Your general point, that clone maps have gaps and genetic maps (and other top down maps) have low resolution and together they produce a reasonable map, is correct,. This provides the rationale for the community involvement, which I tried to make clear. I'll be happy to discuss more if that would help. I also changed the section where the two labs are starting on human. This venture into human was very much an extension of the partnership that we had formed for the worm (and yeast) and I made changes to reflect this. It is absolutely true that the WT was pushing for progress on human and Sanger was ahead in volume. WashU had a lower throughput, but we were sharing fully in dealing with the new problems that human sequence brought. For example, Phil Green adapted his PHRED base calling and PHRAP assembly programs to deal with human during this time.

Reading your account did stimulate me to dig into just when and how we

came to the daily release of assemblies greater than 1 kb. We lack notes on some of the crucial decisions, but here is my best reconstruction. As you point out, we did not start with that. The automated assemblies were problematic (sometimes just bad!) and as newcomers, we undoubtedly felt the pressure to at least meet the standards of the sequencing community -- contiguous, low error rate sequence. By Feb, '94 we had our paper describing 2.2 Mb of contiguous sequence accepted, so probably felt in a stronger position. We had also had experience with multiple clones where everything was basically done except for one or two small problem regions which were recalcitrant to standard solutions. That held up release of otherwise perfectly good sequence. The result was that we began to make those nearly finished clones available (apparently on request at first). Here is what we put in the Feb, '94 WBG:

"Protein similarity data for a part of the region is presented in Table 1 (database hits from the 2.181 Mb sequence are not included in this list). The two groups have used different criteria for determining which cosmids were included in the list. Cambridge has included only finished cosmids and those cosmids which are contiguous but still have one or more problem areas. St. Louis has also included cosmids which have one or two gaps but which are otherwise in good shape. Also, Cambridge has indicated the position and type of similarity within each cosmid, while St. Louis has listed the name and blastx score for the strongest hits. For future submissions we hope to be more consistent; our experience here should help us decide where to set boundaries for future Gazette releases. Although the cosmids which contain database hits may not be complete, the Consortium will make preliminary sequence data available to the community with the caveat that it is preliminary and may still contain errors. Furthermore, we are willing to help locate genes for persons having a bit of sequence data (or to provide an estimated completion time for a particular cosmid)."

We are clearly inviting people to look at the homologies we have found in nearly finished cosmids. But it seems that we are not posting these and instead making them available upon request. We are also inviting inquires about clones earlier in the process.

Some time between there and summer 95 we switched to initial assemblies and posting them on the ftp sites. I presume this is the result of improved initial assemblies (due to PHRAP and PHRED) and a positive experience with our sharing of the sequence with the community. It may have also been triggered by our increasing involvement with human sequencing. I'm only guessing here. Some more notes would be nice. But here is a comment from John after he talked about a chromosome 22 workshop he must have attended in 1994:

"A key step in gaining the trust of the community was to make it an absolute rule that all the sequence produced at the Sanger Centre, whether from worm, yeast or human, would be immediately released into the public domain."

Also from a press release about BRCA2:

Mon Nov 20 21:40 CST 1995

>Large-scale Human Genome Sequencing Aids Cancer Gene Searches
>----------
>
[...]
>both the Washington University Genome Sequencing Centre, St. Louis and the
>Sanger Centre, Cambridge. In accord with our standard practice of
>releasing data to the scientific community as quickly as possible, the DNA
>sequence is being made publicly available by ftp at the 2 centres in
>preliminary form (as assembled shotgun data). In order to aid researchers
>in identifying the genes...

> In order to aid researchers in identifying
> the genes, and in accord with our general policies, this sequence is
> being made publicly available in preliminary form as assembled shotgun
> sequence via our ftp sites (addresses). Importantly, Drs. Wooster and
> Stratton have agreed to immediate release of the data, without delay
> or prior access.

And then from notes from our HUMAN ANALYSIS SUMMIT (January 28-30, 1996), Sanger Centre, UK:

Data on ftp server:
---
*Sequence data should be there:
Data goes to the ftp server weekly (sanger) and nightly (st. louis)
and should for human data be logged so we know the first night it is
available.

Interesting that Sanger is only weekly! Logging the release data presumably relates to patent issues.

Do let me know if you have other questions.

Best,

Bob