

Organizing note and update to the *Agaricus bisporus* genomics community of interest: 3 items of business as the annotation effort draws closer.

Hello, colleagues,

The JGI has now produced a series of scaffolds of the *Agaricus bisporus* H97 haploid genome<sup>1</sup> at about 8.5 x coverage, using Sanger sequencing. There are fewer than 30 scaffolds corresponding to the 13 nuclear and 1 mitochondrial chromosomes<sup>2</sup>. The largest of the "non-large" scaffolds is about 50 kb, so these represent very small areas of uncertainty within the emerging assembly. While the current assembly seems remarkably clean and clear, it is now being further corrected and improved at JGI with the help of a detailed linkage map and known repeated elements that have been provided by the *Agaricus* community.

Once the H97 assembly is perfected, it will then be fed into the machine annotation pipeline at JGI. This involves multiple steps, including training of the gene recognition software using known genes and multiple criteria, and associating putative gene models found in silico with those deposited in global databases and/or provided by collaborators (including several hundred ESTs from *A. bisporus*). Concurrently, we understand, a large number of expressed message cDNAs from three developmental libraries of U1 (comprising both H97 and H39 haploid genomes) will be layered onto the annotation of H97.

When the machine annotation is completed, the data will be put up on the JGI's community sequencing web portal for annotation. Access to those data will be restricted to the community of interest for a period of 6 months, following which they will become public. We must organize a community-wide annotation effort to coincide with that 6-month window of opportunity, following which the organizers of the project are obligated to JGI to prepare a 'master paper' on the nature of the *A. bisporus* genome. We hope that every interested person will be able to make some appropriate contribution to the annotation effort, drawing upon their own local resources. For example, this is potentially a significant training opportunity for able students under capable supervision.

### ***Policy***

The first item of business is policy. We have prepared a brief statement that summarizes our expectations of participants in this effort, and the conditions under which the work will proceed. We think that it is clear, fair, and consistent with JGI's own expectations of us: "Early access to the confidential *Agaricus bisporus* genome sequence data provided by JGI is allowed solely for community-based annotation purposes, with the requirement that there shall be no private exploitation of, or submission for publication of, any aspect of these data until the community annotation effort has achieved sufficient progress to allow preparation of a community-authored 'master report' on the *A. bisporus* genome for submission and publication. This embargo on the use of these data will last at least six months from the time JGI provides the machine-annotated sequence to the community, and possibly longer, depending on the rate of progress of the community-based annotation effort. Annotators granted early access to JGI data are expected to make all reasonable effort toward the completion of their portion of the genome annotation within the initial six month early-access period, and to voluntarily refrain from developing other publications from these data until a master report on *A. bisporus* has been submitted." We hope this (draft) statement will make sense to you. Comments on this policy are welcome. If any technical application of the sequence data is to be made prior to submission of the master report, it should be done in a way that is open to the community.

### ***Annotation***

The uploading of the machine-annotated assembly and JGI's 'six-month restricted access period' may begin in July or August, 2009. We would like to prepare a community paper on our findings in early 2010. That brings us to our second item of business: organizing the annotation work to carry out all of the important tasks without duplication of effort. In this regard, we have prepared a very preliminary table indicating the groups of gene families that must be attended to, and the persons or groups most interested or possibly best suited to take on each section of the annotation work.

We need an active response from each participating member of the community<sup>3</sup>. We hope that

the tentative task assignments will be acceptable to all, but there is room for discussion even though time is getting short. Please make sure that we hear from you as soon as possible. Mike and Rick will coordinate the distribution of necessary tasks, with other input as needed. We intend to have a final work plan locked down and distributed by the end of July 09.

### ***Training***

The last item of business is training. The JGI web portal will provide some very powerful, but somewhat complex, manual annotation tools. JGI is very proactive about training community-based users on these tools. We encourage you to visit the tutorial site entry page at

<http://genome.jgi-psf.org/Tutorial/Tutorial.home.html>

as soon as is convenient. Shortly we will follow up with more information on active training in annotation, possibly as a webinar. Our ambitious goal is to have the Agaricus community prepared for the annotation effort by late summer 2009, to make the best use of the six-month community-access window.

That summarizes where the project stands and where the community effort is headed. We appreciate your interest in this project and will welcome your contributions to the Agaricus bisporus genome project.

--

Rick Kerrigan <[rwk@sylvaninc.com](mailto:rwk@sylvaninc.com)>

Mike Challen <[mike.challen@warwick.ac.uk](mailto:mike.challen@warwick.ac.uk)>

Anton Sonnenberg <[anton.sonnenberg@wur.nl](mailto:anton.sonnenberg@wur.nl)>

### Footnotes:

<sup>1</sup> Sequencing of a second haploid genome (JB137-s8) will proceed shortly at JGI, and should provide a view of diversity within the species.

<sup>2</sup> Occasional updates on the Agaricus Genome Project and other related information are made available via <http://blogs.warwick.ac.uk/mikechallen/>

<sup>3</sup> The attached xls file is a 'living document'. It aims to indicate areas that need coverage and reflects some early expressions of interest. Please feel free to revise (using track changes), indicating and/or confirming your contribution to the project as appropriate, and return the sheet to Mike Challen. As the plan develops we'll try to establish a web-based version.