

Integrating Statistical, Ecological and Evolutionary approaches to Metagenomics (iSEEM)

Project Progress Report

March 31, 2008

1. Progress made towards completion of goals

A. Collaborative wiki set up and being used.

We have set up a wiki site within the Eisen lab wiki for collaboration and communication on the project. It is now being used by all three labs to share information and to keep notes on progress on the project and is helping to link the labs together. Among the things the site is being used for are meeting notes, drafting job advertisements, and drafting papers. We are using the MediaWiki software for the site and overall it has been a useful tool for communication. One limitation is that the discussion-forum functions are less than ideal within MediaWiki and we are working to install a better system for tracking discussions. One good feature of this system is that any page within the wiki can be moved from being private (as they are now) to being public. Thus we can collaboratively create sites for releasing information to the broader community and then make them public very easily.

In addition, we are working on a possible blog for the project where the PIs and other personnel will write about metagenomics and our work for a broader audience. We have reserved the domain name <http://www.iSEEM.org/> for this blog and/or for a project web site.

B. Review paper on “Computational Methods for Studying Microbial Diversity”

We are jointly working on a review paper on “Computational Methods for Studying Microbial Diversity” for the journal PLoS Computational Biology. We are using this to get everyone in the project up to speed on what has been done in the broader scientific community in the area of the goals of our project. In addition, the private web page we are making within the Wiki for writing this paper has links to web servers, available software, and journal publications in this field. We will make this web resource available to the community through our web sites, through PLoS Computational Biology and through CAMERA.

C. Hiring personnel

In January 2008 (when the money was first approved at Davis) we began recruitment of five postdocs and a bioinformatics engineer. We composed a job advertisement that describes the iSEEM project, principal investigators, qualifications for both types of position, and the recruitment process. In February and March 2008, this ad was distributed to our personal contacts, posted on lab and university websites, and placed on sciencecareers.org (for 8 weeks). To collect application materials, we designed an

application “portal” on the UC Davis Genome Center website (http://jobs.genomecenter.ucdavis.edu/start_app.php?job_id=77 and http://jobs.genomecenter.ucdavis.edu/start_app.php?job_id=78). The portal includes a feature that allows the PIs to review materials, enter comments, and score applicants on criteria specific to the project.

To date we have received 39 postdoc applications and 15 bioinformatics engineer applications. In March 2008 we conducted a preliminary review of applications. From this review, we produced a short list of 12 postdocs and 4 engineers. We plan to conduct phone interviews with short list candidates. A few candidates will be interviewed in person by one or more PIs, for example, if we are at a conference or other event with them in the coming months. Two postdoc candidates will be in Davis in the next month, and will present talks to the Eisen and Pollard labs. Our goal is to make job offers by May 2008 with start dates in Summer 2008. As we conduct interviews, we will continue to review applications until the positions are filled.

D. Scientific research

- Dongying Wu in Eisen lab has been working on rRNA alignment and OTU identification. We have completed a software tool that can generate multiple sequence alignments of rRNA genes automatically and are nearly finished with a tool that will identify operational taxonomic units from the alignments for diversity assays.
- Dongying Wu in Eisen lab has initiated automated protein family clustering work and is designing scripts for identifying families using Lek clustering.
- We have initiated discussions with Kimmen Sjolander at U. C. Berkeley about using the protein families she will identify through her new NSF funded “Bacterial phylogenomics” project (Eisen is a CoPI on her grant). She has agreed and in the long run we will probably end up using her families as (1) she is perhaps the world’s expert in automated methods to do this and (2) it will save us resources to focus on other work. Temporarily we will still build our own families as the one from the Sjolander lab will not be ready for some months.
- Martin Wu in Eisen lab has been working on automated protein phylogeny tools and identifying the set of 50 protein families to use for our initial analyses.
- Jessica Green is leading an initial analysis of distance-decay relationships in rRNA data using the GOS data.

2. Quarterly Meeting

Technically we have not yet had a formal quarterly meeting for the project because we have been waiting to have one during a visit to CAMERA. We have a tentative meeting scheduled at CAMERA for April 18, 2008 where we will both meet with CAMERA personnel and have our 1st quarterly meeting.

We have however commenced regular conferencing via the web. We have held three conferences between the PIs, two of which were video chats and one was a phone

conference call. We plan to have weekly conference calls to keep communications open. Notes from these meetings are attached as PDFs.

3. Any unexpected challenges that imperil successful completion of the Outcome.

The only unexpected challenge so far was the time delay in getting Davis to sign off on the grant. After some initial delays, we were given the go ahead to start spending on January 28, 2008.