

KNIGHT - QIIME

A. Project Status - Comparison to Metrics for Year 1

1. Build initial links between QIIME and MG-RAST and VAMPS.

We have achieved this via the “QIIME report” functionality in MG-RAST, which produces QIIME-format tables, and by the decision that the topology of the data exchange will be that QIIME and VAMPS will exchange data via MG-RAST rather than exchanging directly with each other, greatly simplifying the necessary interfaces. Bi-directional exchange of information between QIIME and MG-RAST has already been achieved, resulting in the deposition of thousands of public QIIME samples and metadata in MG-RAST. We have been working closely with the MG-RAST team on implementing and iterating the API. Additionally, we participated in the MBL STAMPS meeting, during which several of the key QIIME developers met with representatives from both Argonne and MBL to further define plans. The QIIME metadata portal is available for all three sites to use.

2. Define interchange formats for OTU tables and related files.

We have defined these formats and implemented them in QIIME for both dense and sparse matrices. Sparse matrix support and additional documentation will be distributed throughout the MoBEDAC sites by Dec 31. Importantly, we are generalizing this format to accommodate all abundance and frequency tables generally, not just OTU tables.

3. Finalize MIENS-compliant I/O and data exchange among sites.

We have done this through the QIIME metadata portal, which allows exchange of tab-delimited text metadata files. The ANL team is also exploring the use of GCDML as an auxiliary format for metadata exchange. Note that the MIENS standard was renamed MIMARKS, and documentation throughout the project will reflect this new name. We agreed at the MBL STAMPS meeting that we would exchange 16S rRNA and metagenomic sequence data files in the standard FASTA format. Data exchange for other types of files, e.g. proteomics, remains to be investigated after additional information is gathered from Sloan grantees.

B. Training of Students and Post-doctoral Researchers

Four students, Antonio Gonzalez-Pena, Dan Knights, Daniel McDonald, and Luke Ursell have been partially supported by MoBEDAC funds and trained on and/or contributed to the development of MoBEDAC tools. They will be expected to be authors on a paper describing the QIIME/MG-RAST integration effort and its application to large datasets such as the HMP dataset.

Three postdocs, Bharath Prithiviraj, Jose Clemente-Litran and Jesse Stombaugh, have contributed to MoBEDAC tool development and have been partially supported by the grant. Bharath has primarily been contributing his expertise on biogeography and biomarkers, Jose his expertise on taxonomy assignment, and Jesse has been leading

the database effort at our end (with assistance from technician Doug Wendel, who has primarily been working on metadata).

In addition, Knight has presented tutorials at the International Human Microbiome Consortium meeting in March 2011, the Hopkins Marine Microbiology Course and the Ft. Collins Comparative Genomics Workshop in July 2011 and at the STAMPS course at MBL August 2011, thus providing additional student training.

C. Publications:

- Chicago and Colorado have 1 publication on MG-RAST to be submitted to a peer-reviewed journal by the end of the year.
- We anticipate a publication on the BE metadata package being submitted in the next 3 months.

D. Presentations/Talks:

All of these talks covered aspects of the MoBEDAC integration effort, among other topics, and credited Sloan funding:

- 2011 Gordon Research Conference (GRC) Applied & Environmental Microbiology, South Hadley, MA, July 10-15. Introduction to the session Evolution of Bacterial Species and their Genomes. Oral, invited speaker and discussion leader.
- 2011 International Society for Exercise Immunology (ISEI), Oxford, UK, July 11. "Diet, genetics, and the mammalian microbiome". Oral, invited speaker, keynote speaker.
- 2011 The First International EMP Workshop, Shenzhen, China, Jun 13-15. "Lessons learned from Massive Datasets". Oral, invited speaker, session chair.
- 2011 Marabou Nutrition Symposium, Stockholm, Sweden, Jun 10-12. "Defining the human microbiome". Oral, invited.
- 2011 Conference on Bacterial Genetics and Ecology (BAGECO11), Corfu, Greece, May 29 – Jun 2. "Spatial and temporal patterning of the human microbiome". Oral, invited speaker.
- 2011 Keystone Symposia, Breckenridge, CO, Mar 25-29. "Quantitative Insights into Microbial Ecology". Oral, invited speaker, plenary session chair.
- 2011 JGI Genomics of Energy and Environment, San Francisco, CA, Mar 22-24. "Spatially and Temporally Resolved Studies of the Human Microbiome". Oral, invited speaker.
- 2011 Annual R.G.E. Murray Lectureship, University of Western Ontario, London, Canada, Mar 21. "Variation in the human microbiome across space in time". Oral, invited speaker.
- 2011 International Human Microbiome Congress (IHMC), Vancouver, Canada, Mar 9-11. "I Believe The Robots Are Our Future: Automation, Machine Learning, and 10,000 Samples Per Study". Oral, invited.

2011 Advances in Genome Biology and Technology (AGBT) meeting, Marco Island, FL, Feb 2-5. "Spatially and temporally explicit studies of the human microbiome". Oral, invited, plenary speaker.

E. microBE.net

In future we will provide microbe.net with information regarding MoBEDAC presence at upcoming meetings and conferences as well as, newsworthy MoBEDAC accomplishments.

F. Lab contact person

Gail Ackermann, our lab metadata specialist, will be partially supported from Sloan funds going forward. Her contact details are:

Gail Ackermann
Gail.ackerman@colorado.edu
303-735-1664
Department of Chemistry & Biochemistry
UCB 215
University of Colorado at Boulder
Boulder, CO 80309